Insper

Fundamentos de aprendizagem estatística

Aula 1

Magno Severino
PADS - Modelos Preditivos
23/04/2021

Referências bibliográficas do curso

- An Introduction to Statistical Learning: with Applications in R. James, G. and Witten, D. and Hastie, T. and Tibshirani, R. 2013.
- The Elements of Statistical Learning. Hastie, T. and Tibshirani, R. and Friedman, J. 2017.
- R for Data Science Wickham, H. and Grolemund, G. 2017.
- Aprendizado de Máquina, uma abordagem estatística. Izbicki, R. and Santos, T. 2020

Critério de avaliação

- Quiz: 10%
- Listas: 20%
- Atividade integradora: 70%
- Seminário (extra): até 5%

Objetivos de aprendizagem

Ao final dessa aula você deverá ser capaz de

- Diferenciar um problema de regressão e um de classificação.
- Conceituar o que é um modelo estatístico.
- Definir os dois objetivos de modelagem (predição e inferência).
- Compreender o *trade-off* entre interpretabilidade e flexibilidade de um modelo.
- Compreender como mensurar o erro de um modelo.
- Conceituar a decomposição do erro em viés e variância.

Fundamentos de aprendizagem estatística

Modelos supervisionados	Modelos não supervisionados

Fundamentos de aprendizagem estatística

Modelos supervisionados

- Regressão linear
- Regressão Logística
- Regularização
- Árvore de Classificação
- Árvore de Regressão
- Bootstrap e Bagging
- Floresta Aleatória
- Boosting

Modelos não supervisionados

- Redução de dimensão
- Escalonamento multidimensional
- Análise de agrupamento
- K-médias
- Text mining

Fundamentos de aprendizagem estatística

Modelos supervisionados

- Regressão linear
- Regressão Logística
- Regularização
- Árvore de Classificação
- Árvore de Regressão
- Bootstrap e Bagging
- Floresta Aleatória
- Boosting

Modelos preditivos

Modelos não supervisionados

- Redução de dimensão
- Escalonamento multidimensional
- Análise de agrupamento
- K-médias
- Text mining

Modelagem preditiva avançada

Fundamentos de aprendizagem estatística

Modelos supervisionados

- Problemas de regressão
- Problemas de classificação

Fundamentos de aprendizagem estatística

Modelos supervisionados

- Problemas de regressão

Idade	Anos de escolaridade	Salário
28	13	R$ 6357,70
33	15	R$ 7035,38
45	11	R$ 9853,98
⋮	⋮	⋮
40	9	R$ 1150,76

- Problemas de classificação

Fundamentos de aprendizagem estatística

Modelos supervisionados

- Problemas de regressão
- Problemas de classificação

Idade	Anos de escolaridade	Salário
28	13	R$ 6357,70
33	15	R$ 7035,38
45	11	R$ 9853,98
⋮	⋮	⋮
40	9	R$ 1150,76

Fundamentos de aprendizagem estatística

Modelos supervisionados

- Problemas de regressão

Idade	Anos de escolaridade	Salário
28	13	R$ 6357,70
33	15	R$ 7035,38
45	11	R$ 9853,98
⋮	⋮	⋮
40	9	R$ 1150,76

- Problemas de classificação

Idade	Salário	Tem graduação
28	R$ 6357,70	Não
33	R$ 7035,38	Sim
45	R$ 9853,98	Sim
⋮	⋮	⋮
40	R$ 1150,76	Não

Fundamentos de aprendizagem estatística

Modelos supervisionados

- Problemas de regressão

Idade	Anos de escolaridade	Salário
28	13	R$ 6357,70
33	15	R$ 7035,38
45	11	R$ 9853,98
⋮	⋮	⋮
40	9	R$ 1150,76

- Problemas de classificação

Idade	Salário	Tem graduação
28	R$ 6357,70	Não
33	R$ 7035,38	Sim
45	R$ 9853,98	Sim
⋮	⋮	⋮
40	R$ 1150,76	Não

Fundamentos de aprendizagem estatística

Modelos supervisionados

- Problemas de regressão

Idade	Anos de escolaridade	Salário
28	13	R$ 6357,70
33	15	R$ 7035,38
45	11	R$ 9853,98
⋮	⋮	⋮
40	9	R$ 1150,76

- Problemas de classificação

Idade	Salário	Tem graduação
28	R$ 6357,70	Não
33	R$ 7035,38	Sim
45	R$ 9853,98	Sim
⋮	⋮	⋮
40	R$ 1150,76	Não

Regularização

Regressão linear

Bootstrap e Bagging

Regressão Logística

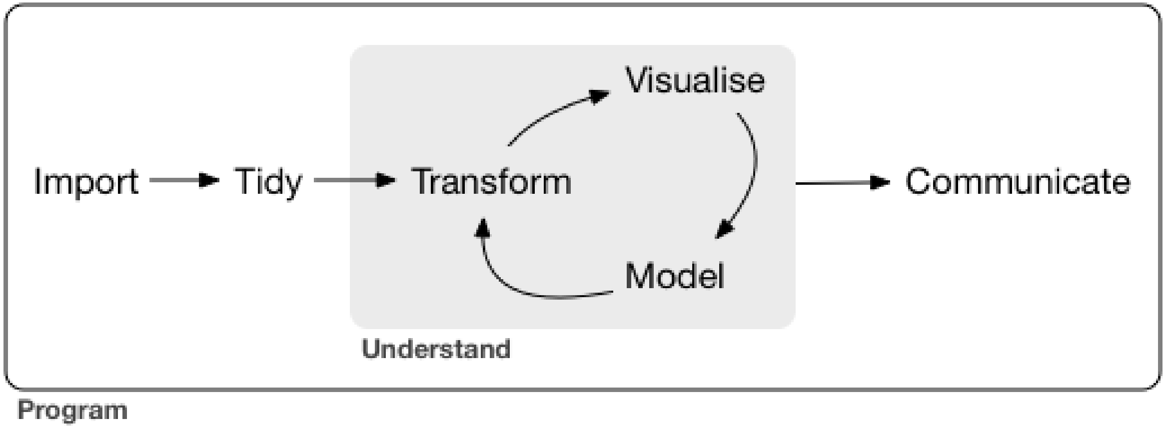
Árvore de Regressão

Floresta Aleatória

Árvore de Classificação

Boosting

Fundamentos de aprendizagem estatística



[1] Fonte: livro R for Data Science Wickham, H. and Grolemund, G. 2017.

Caso: *advertising* [1]

[1] Fonte: livro An Introduction to Statistical Learning with Applications in R.

Caso: *advertising* [1]

Como cientista de dados, você deve fornecer ideias sobre como estimar as vendas de um determinado produto baseado no total investido em mídia (jornal, rádio e TV).

Para isso, foram coletados o total de vendas e o valor investido em cada uma das mídias consideras em 200 mercados diferentes.

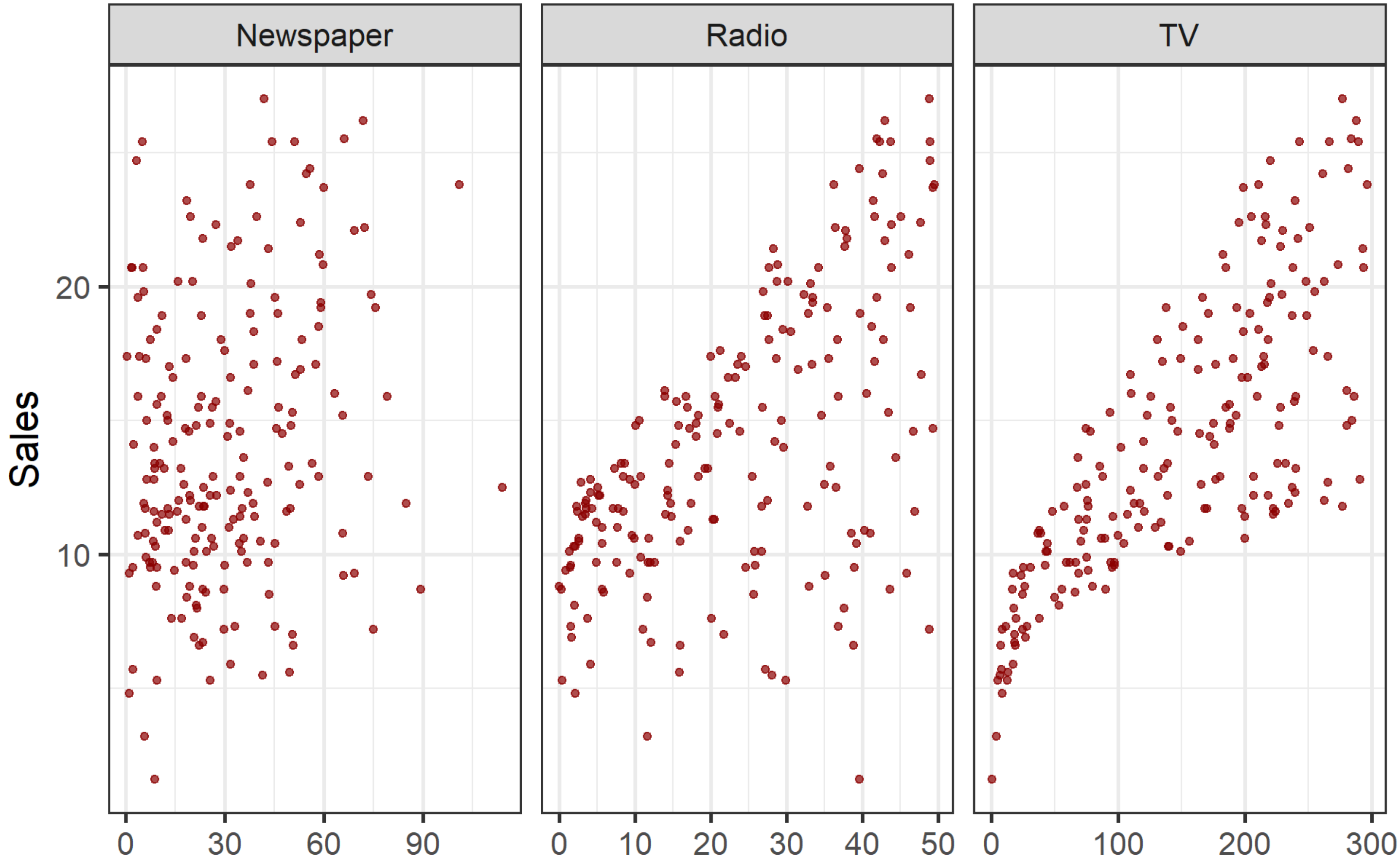
	TV	radio	newspaper	sales
1	230.1	37.8	69.2	22.1
2	44.5	39.3	45.1	10.4
3	17.2	45.9	69.3	9.3
4	151.5	41.3	58.5	18.5
5	180.8	10.8	58.4	12.9

Showing 1 to 5 of 200 entries

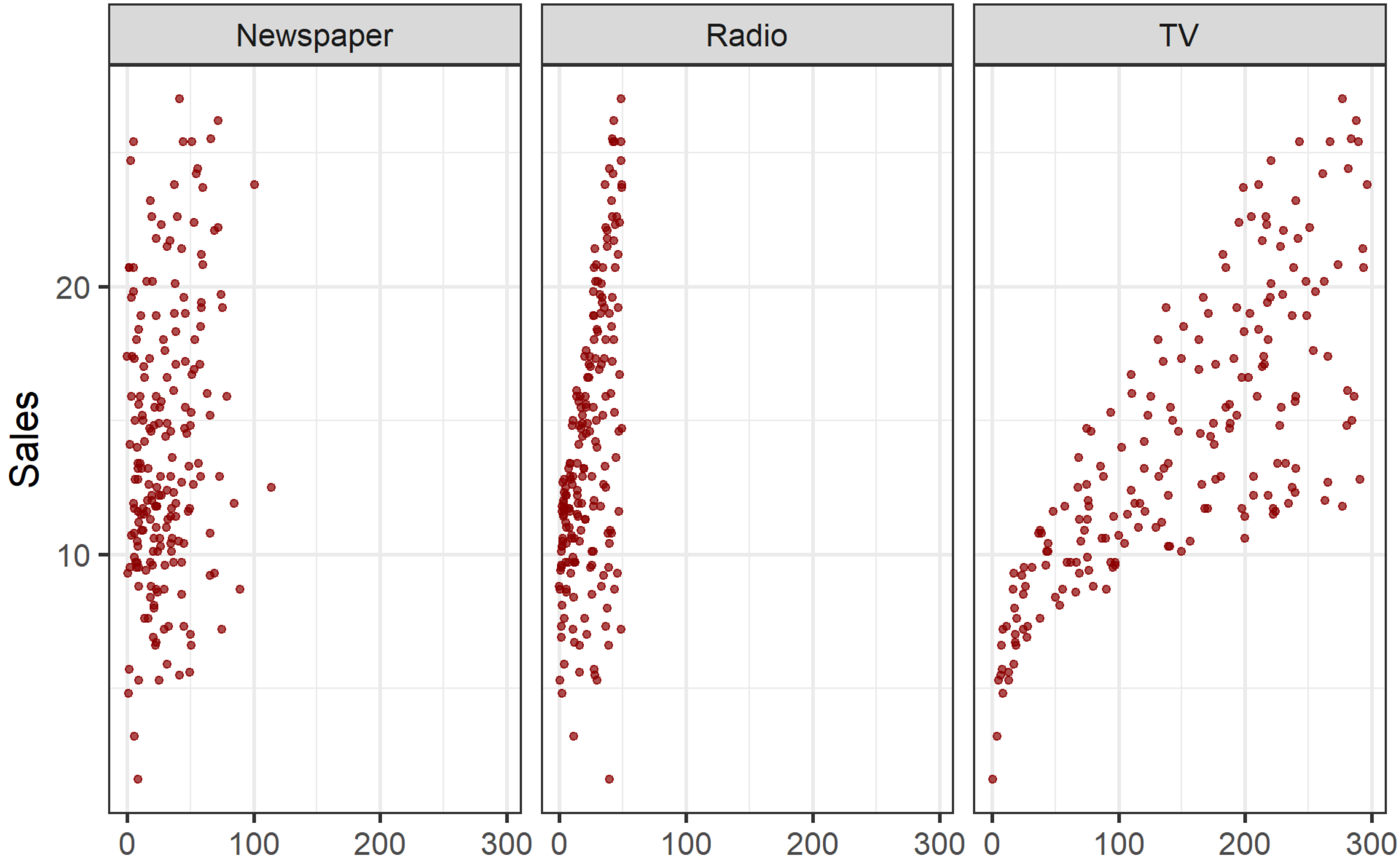
Previous 1 2 3 4 5 … 40 Next

[1] Fonte: livro An Introduction to Statistical Learning with Applications in R.

Visualização dos dados



Visualização dos dados



Um pouco de nomenclatura e notação

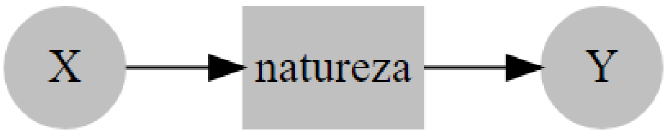
- **`sales`**: variável resposta, variável dependente, *target*, *output*.
  - Geralmente é denotada por Y.
- **`newspaper, radio, TV`**: variáveis explicativas, variáveis independentes, variáveis preditoras, *inputs*, *features*.
  - Geralmente são denotadas por X, com um subscrito para distinguí-las.
  - Por exemplo, X_1 para `newspaper`, X_2 para `radio` e X_3 para `TV`.

Estudando a relação entre variáveis

Assumimos que existe uma relação entre as preditoras, X_1, X_2, X_3 (jornal, rádio e TV), e a resposta que pode ser escrita da seguinte forma

$$Y = f(X_1, X_2, X_3) + \epsilon,$$

em que f é uma função desconhecida das preditoras e ϵ é um erro aleatório que é independente de $X = (X_1, X_2, X_3)$ com média zero.



Quando Y é uma variável quantitativa, dizemos que estamos tratando de um problema de **regressão**.

Notação

Observação	newspaper	radio	TV	sales
1	69,2	37,8	230,0	22,1
2	45,1	39,3	44,5	10,4
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
200	8,7	8,6	232,0	13,4

Observação	X_1	X_2	X_3	Y
1	X_{11}	X_{12}	X_{13}	Y_1
2	X_{21}	X_{22}	X_{23}	Y_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
200	$X_{200;1}$	$X_{200;2}$	$X_{200;3}$	Y_{200}

Notação geral

Observação	X_1	X_2	$\cdots$	X_p	Y
1	X_{11}	X_{12}	$\cdots$	X_{1p}	Y_1
2	X_{21}	X_{22}	$\cdots$	X_{2p}	Y_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	X_{n1}	X_{n2}	$\cdots$	X_{np}	Y_n

O conjunto das p preditoras pode ser representado por $X = (X_1, X_2, \ldots, X_p)$.

X_j é a preditora j.

X_{ij} é o valor da preditora j na i-ésima observação.

Generalizando

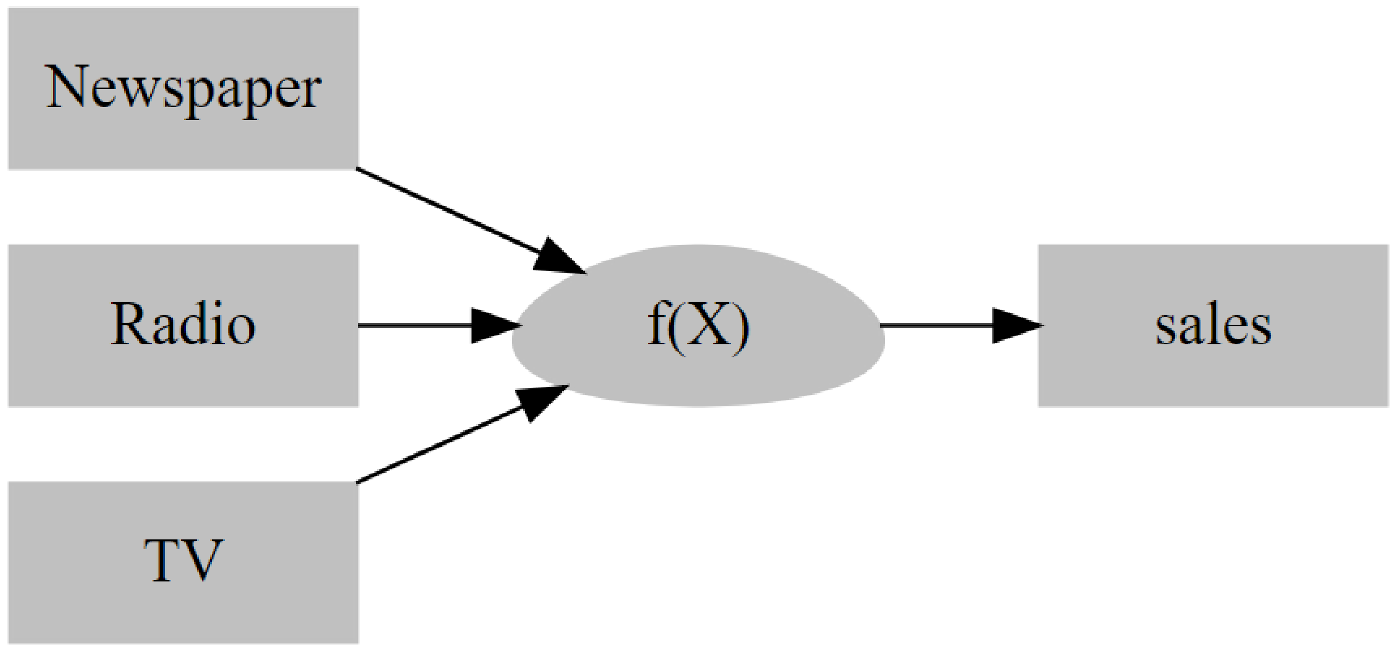
Considere que seja observada uma variável quantitativa Y e p preditoras $X = (X_1, X_2, \ldots, X_p)$.

Assumimos que existe uma relação entre essas medidas que pode ser escrita de forma geral como

$$Y = f(X) + \epsilon,$$

Note que f representa a informação sistemática que X fornece de Y.

Novamente, ϵ é um erro aleatório que é independente de X com média zero. Ele representa fatores que afetam Y, mas não estão relacionados aos valores das preditoras X.


Newspaper
Radio
TV
f(X)
sales

Porque estimar f?

Existem duas principais razões: *predição* e *inferência*.

Inferência

O interesse está centrado em entender como Y varia de acordo com os valores de $(X_1, \ldots, X_p)$:

- Quais variáveis (X_j) estão associadas com a resposta (Y)?
- Qual a relação entre a resposta (Y) e cada preditora?
- A relação entre Y e cada preditora pode ser representada adequadamente usando uma equação linear, ou é uma relação mais complicada?

Caso: *advertising*

- Qual mídia contribui para o total vendas?
- Qual mídia gera o maior aumento no total vendas?
- Qual o aumento esperado no total de vendas dado um acrescimo no investimento de propaganda em TV?

Predição

- Em algumas situações, queremos apenas estimar o valor de Y dadas as variáveis preditoras $(X_1, \ldots, X_p)$.

- Estimativa para f: $\hat{f}$.

- Nesse caso, trata-se a função preditora $\hat{f}$ como uma caixa preta.

- Não há interesse particular na forma de $\hat{f}$, desde que ela gere boas predições para Y.

- **Este é o foco deste curso.**

Caso: *advertising*

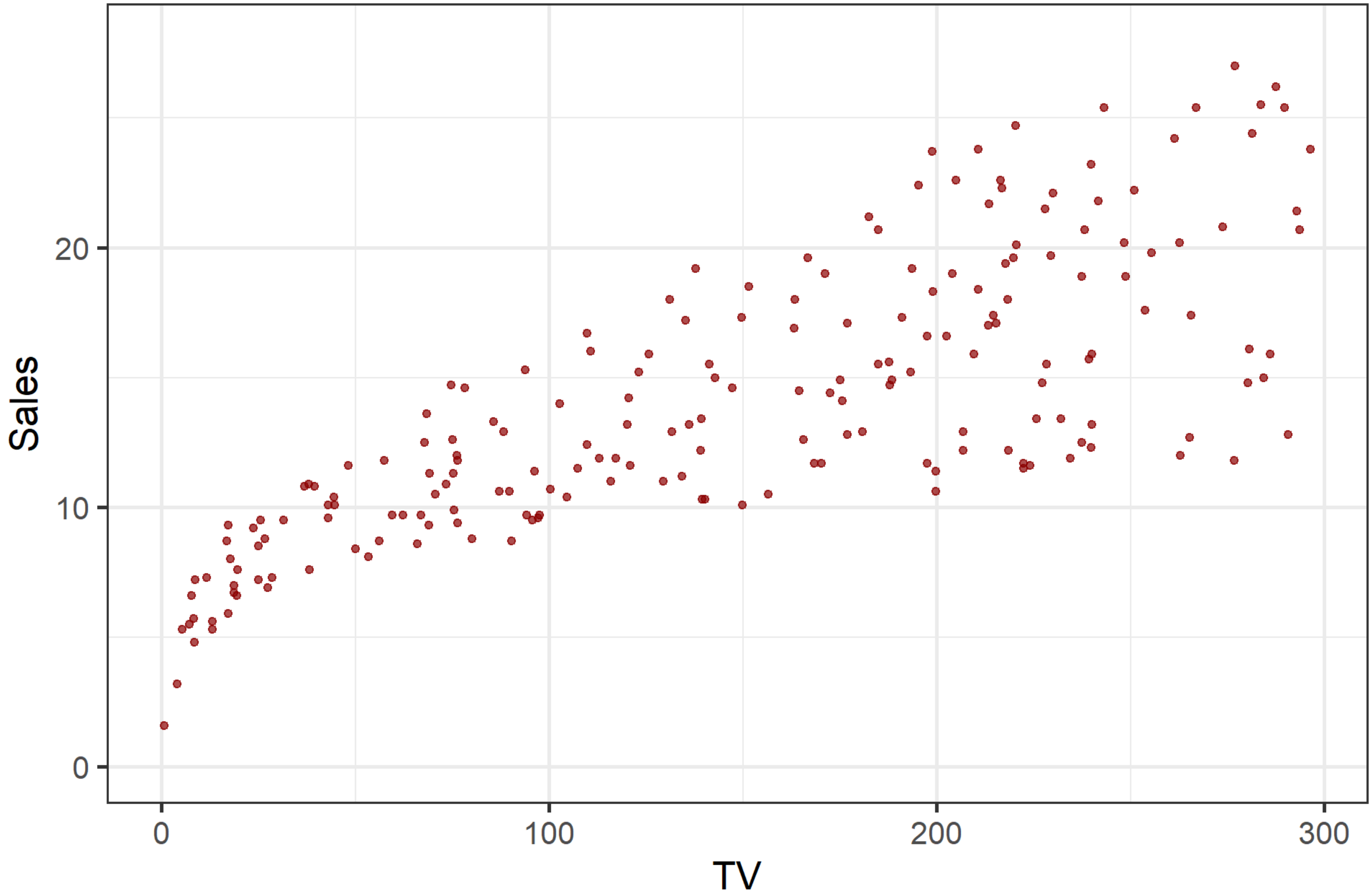
- Qual o total de vendas previsto para um produto em que foi investido 32,0 em `newspaper`, 24,3 em `radio` e 142,5 em `TV`?

Como estimar f?

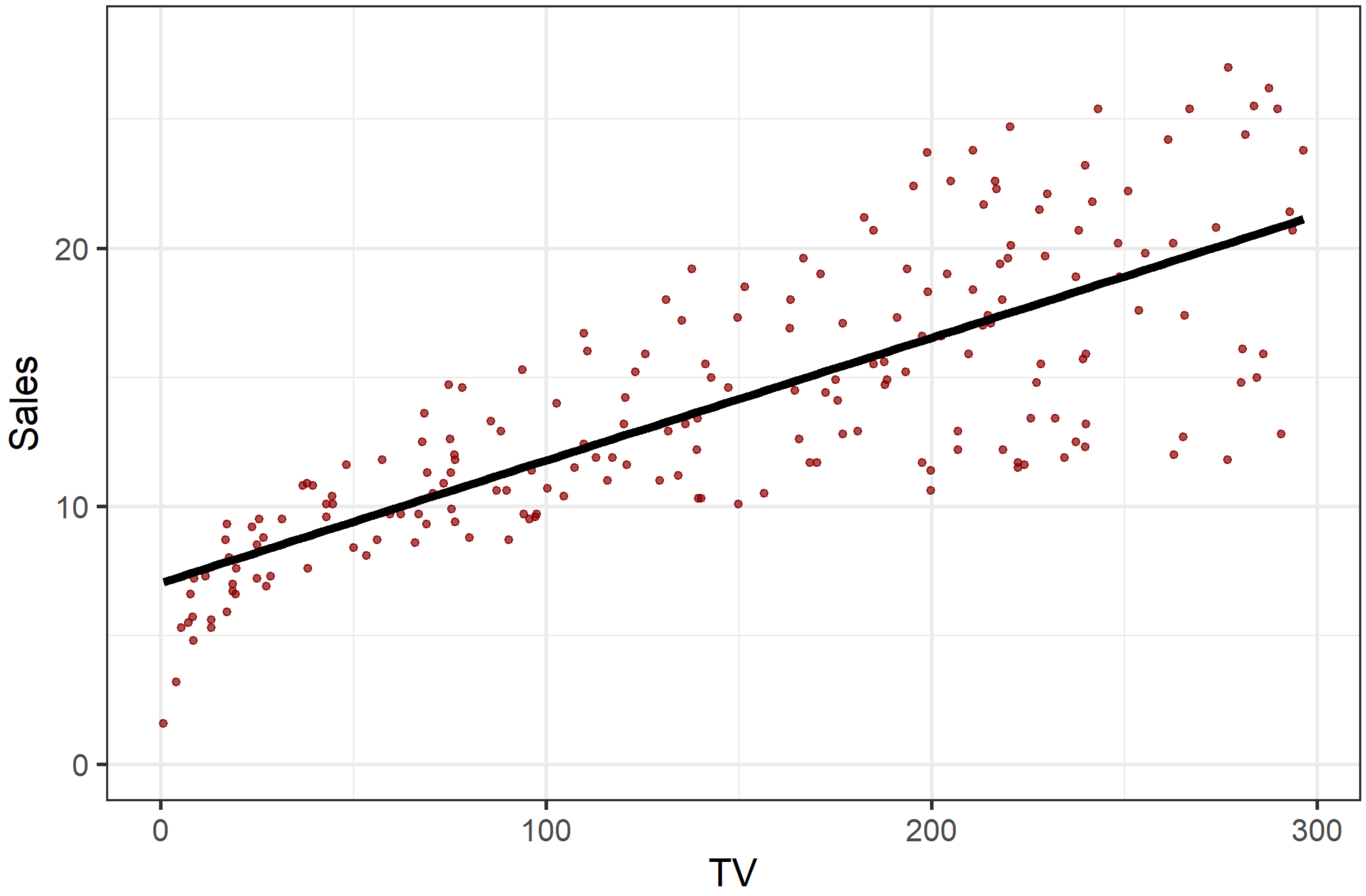
- Considere que foi observado um conjunto de dados tamanho n.
- Essas observações são chamadas de **conjunto de treino/treinamento**, pois serão usadas para treinar o modelo para estimar f.
- Seja $x_i = (x_{i1}, \ldots, x_{ip})$ o vetor que presenta os valores das p preditoras na oservação i.
- Dados de treino: $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$.
- No caso *advertising*, $n = 200$ e $p = 3$.
- Aprendizagem **supervisionada** (quando Y é conhecido).
- Ideia: utilizar a informação contida nos dados de treinamento para construir uma função $\hat{f}$, que será nossa estimativa para f, e assim prever

$$\hat{Y} = \hat{f}(X).$$

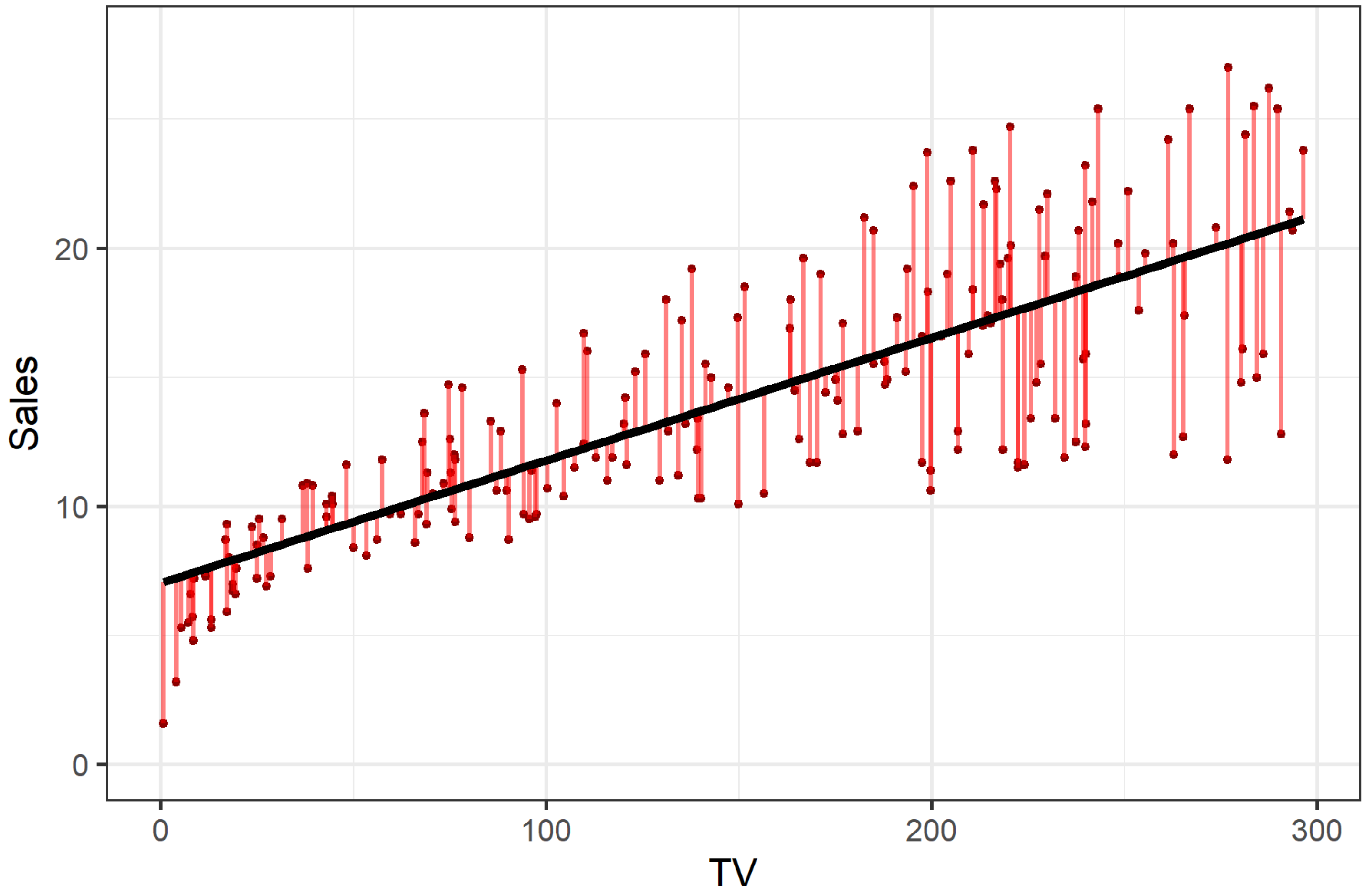
Visualização dos dados



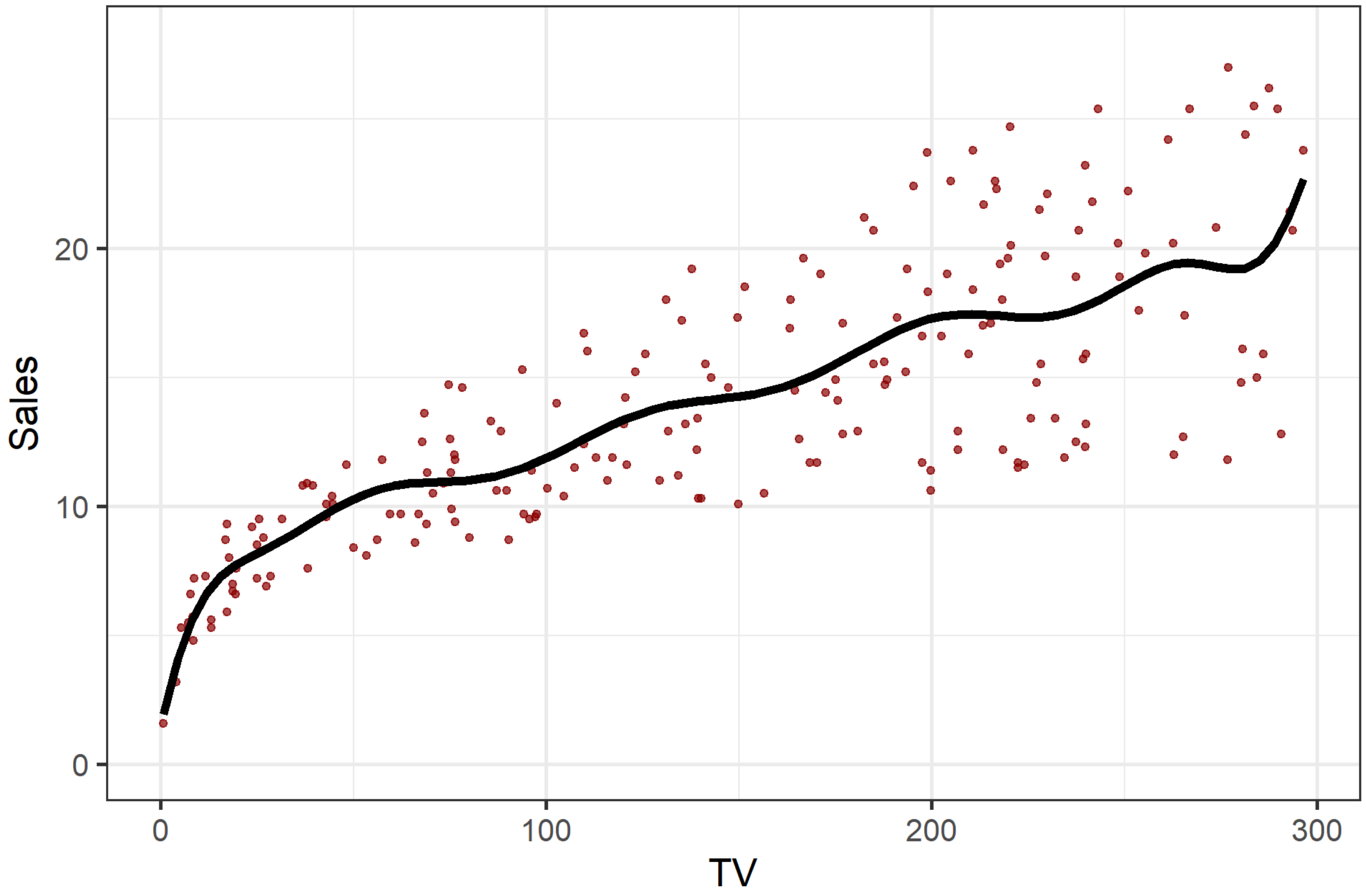
Modelo linear



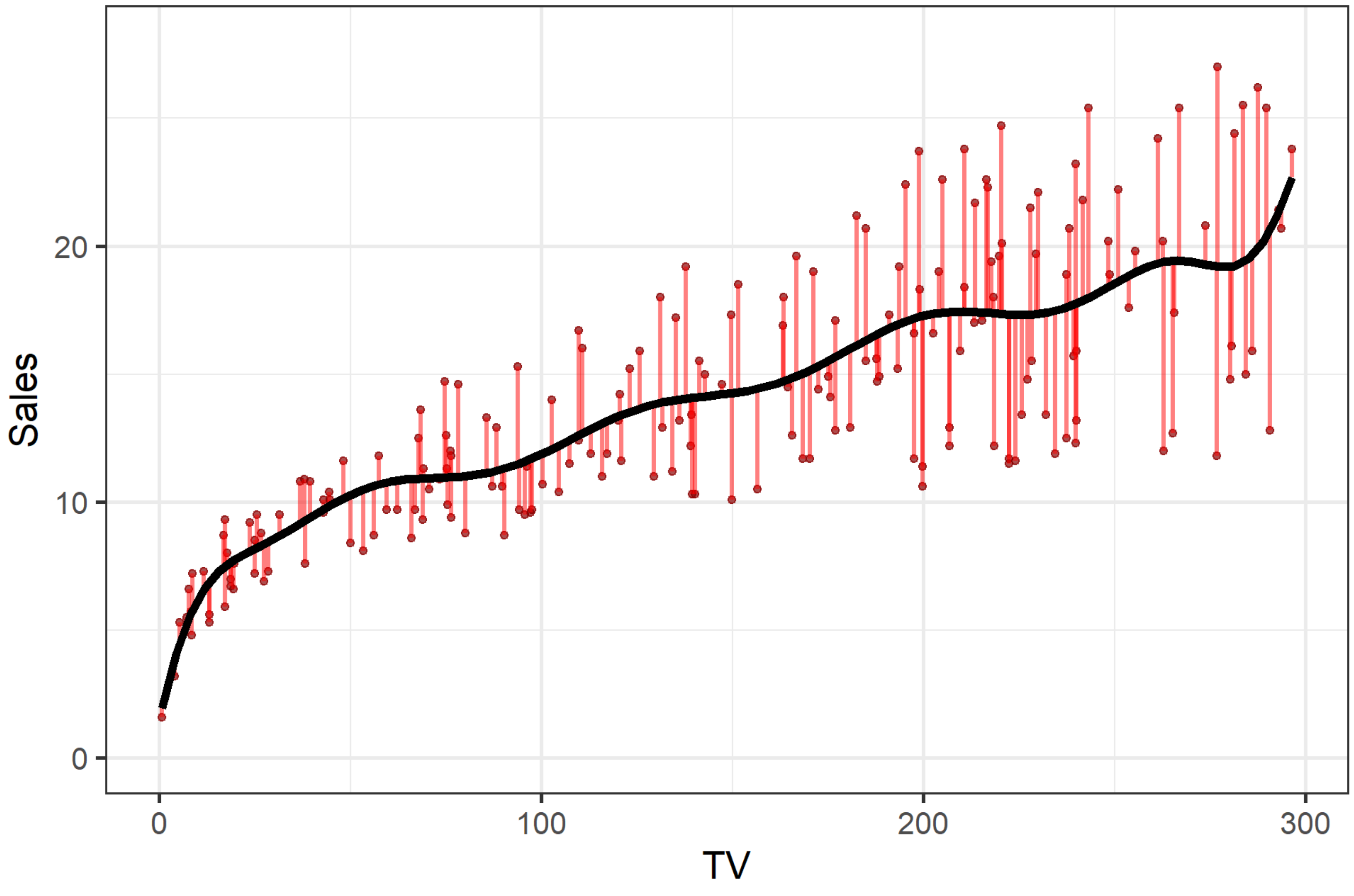
Modelo linear



Modelo polinomial



Modelo polinomial



Erro redutível e irredutível

- Em geral, $\hat{f}$ não é uma estimativa perfeita para f, essa "imperfeição" produz um erro:

$$Y - \hat{Y} \neq 0.$$

- Esse erro é *redutível*, pois pode-se melhorar a acurácia de $\hat{f}$ utilizando um modelo mais apropriado para aprender sobre f.

- Lembre que

$$Y = f(X) + \epsilon,$$

  ou seja, Y também é uma função de ϵ, que não pode ser predito através de X.

- A variabilidade associada ao ϵ afeta a acurácia das predições.

- Esse erro é conhecido como erro *irredutível*.

- Considere uma $\hat{f}$ e um conjunto de preditoras X. Se fixarmos $\hat{f}$ e X, então

$$\begin{aligned} E(Y - \hat{Y})^2 &= E(f(X) + \epsilon - \hat{f}(X))^2 \\ &= \underbrace{[f(X) - \hat{f}(X)]^2}_{\text{redutível}} + \underbrace{\mathrm{Var}(\epsilon)}_{\text{irredutível}}. \end{aligned}$$

Formas de estimar f

Modelos paramétricos

1. Assume-se a forma de f. Por exemplo, que f é linear:

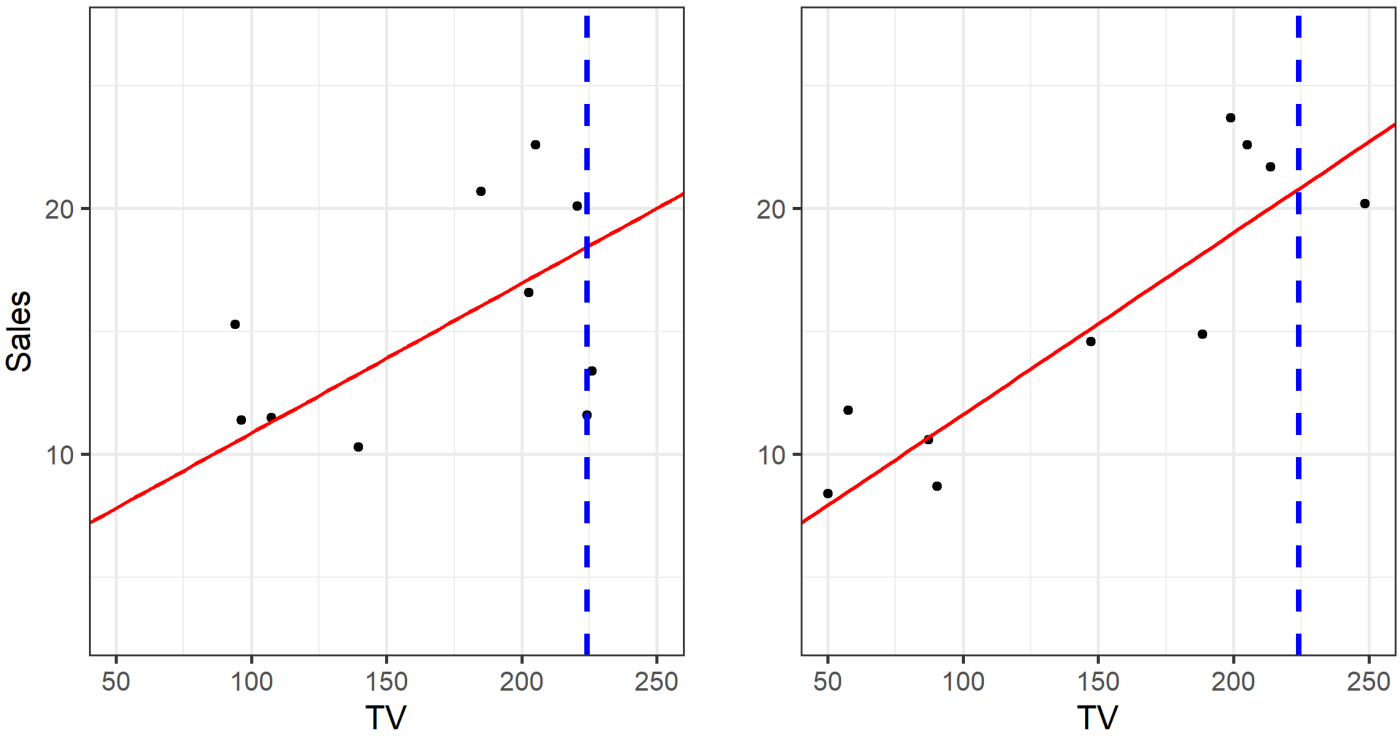
$$f(X) = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p$$

2. Agora o problema é estimar os coeficientes $\beta_0, \beta_1, \ldots, \beta_p$ a partir do conjunto de treinamento.

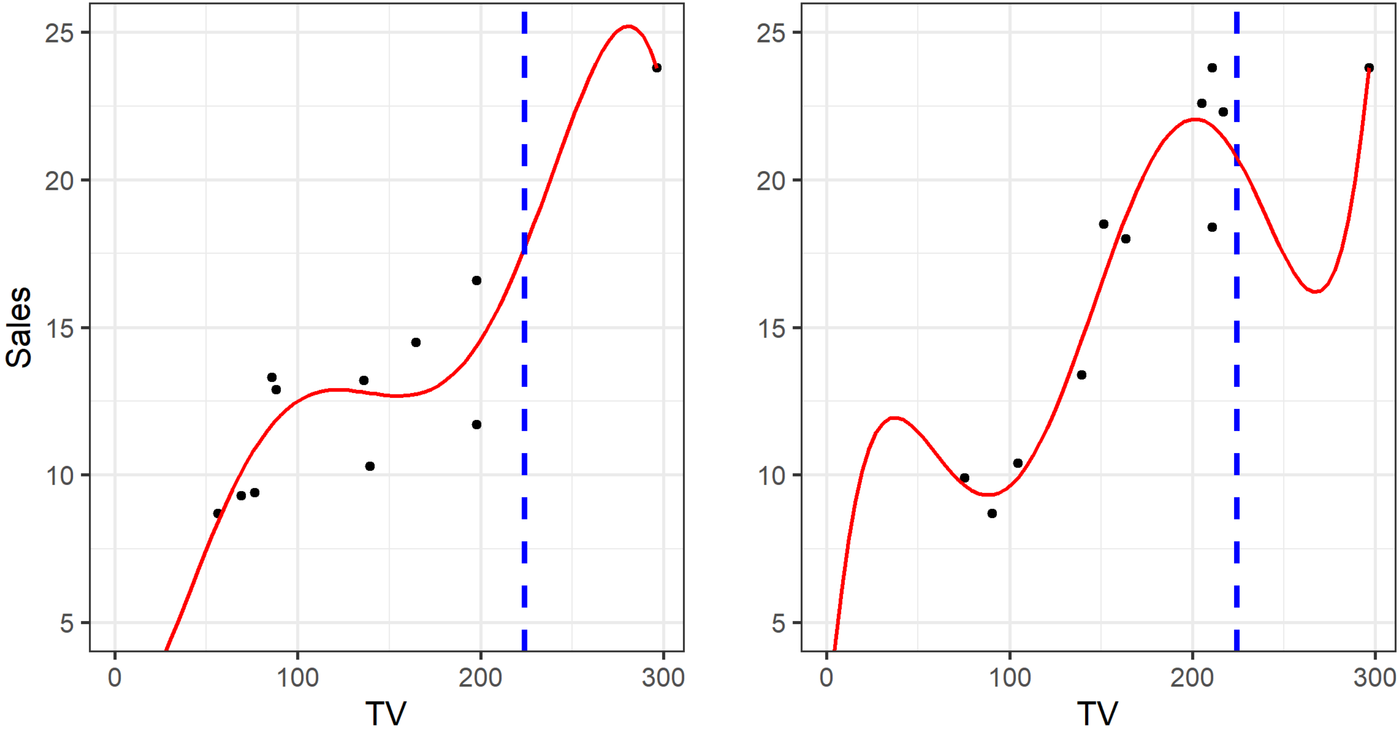
Modelos não-paramétricos

- Não assume nenhuma forma para a função f.
- Busca uma função que seja o mais próximo possível dos pontos no conjunto de treinamento.

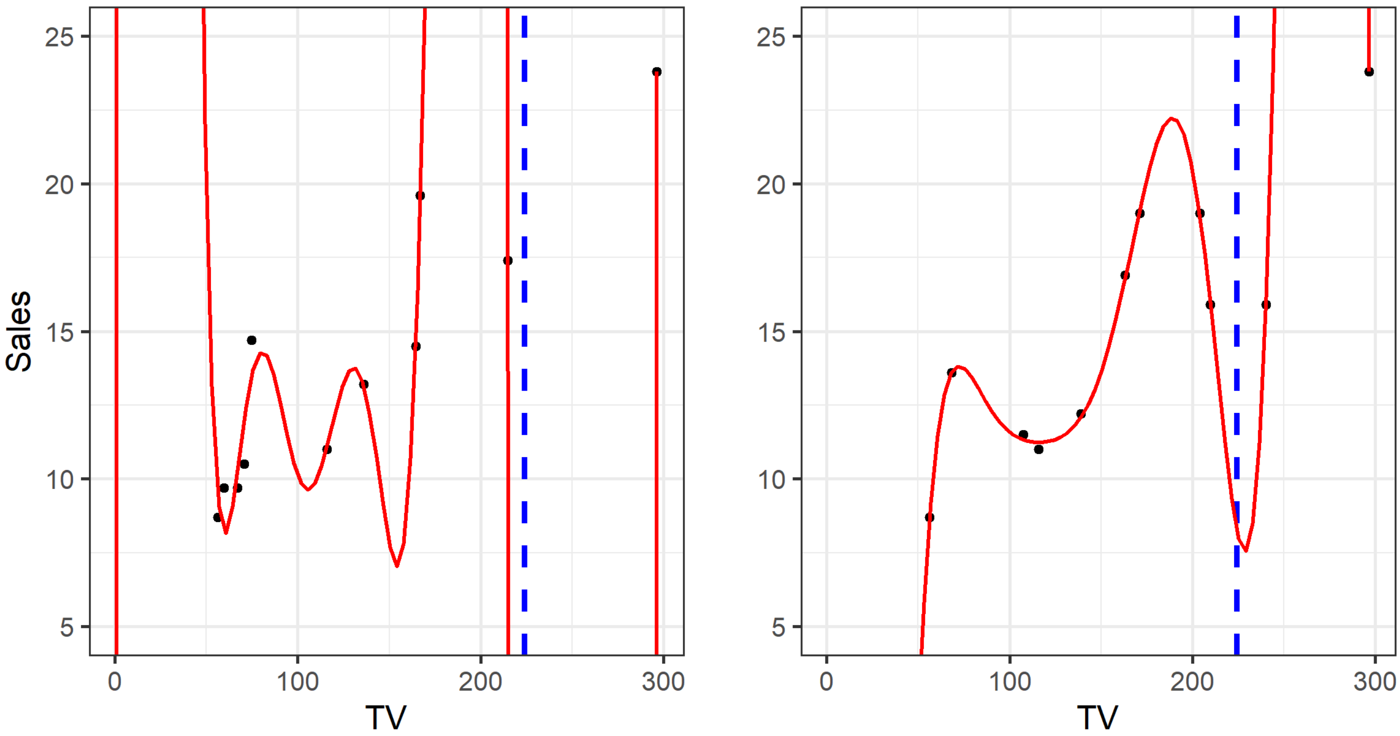
Modelo menos flexível



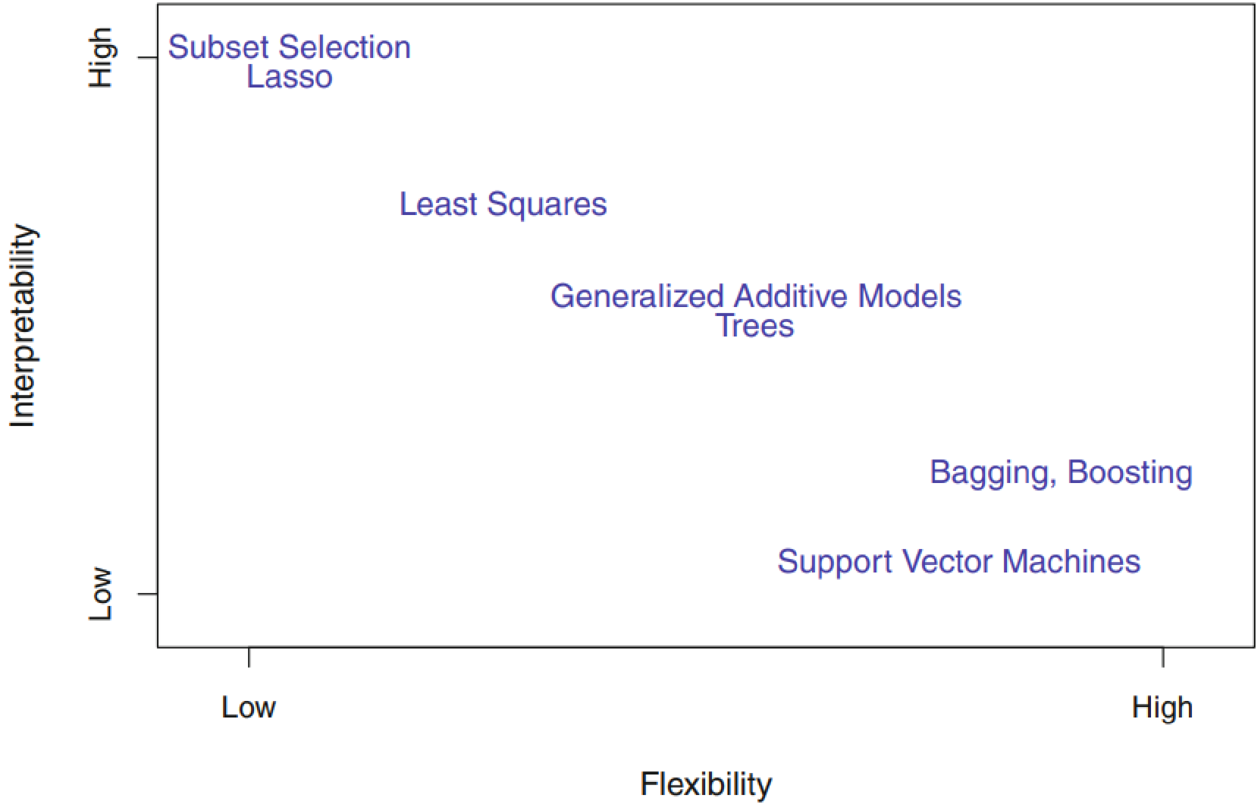
Modelo mais flexível



Modelo muito flexível



Relação entre interpretabilidade e flexibilidade



[1] Figura: pág. 25 de An Introduction to Statistical Learning with Applications in R.

Resumindo...

- Definição de variável resposta Y e variáveis preditoras $X = (X_1, \ldots, X_p)$.
- Quando Y é quantitativa, dizemos estar tratando de um problema de **regressão**.
- Modelo: $Y = f(X) + \epsilon$, f é desconhecida.
- Queremos estimar f para fazer **inferência** ou **predição**.
- Estimativa de f: $\hat{f}$.
- O erro de $\hat{f}$ pode ser decomposto em
  - erro **redutível**: $[f(X) - \hat{f}(X)]^2$,
  - erro **irredutível**: $\mathrm{Var}(\epsilon)$.
- Relação entre interpretabilidade e flexibilidade.

Proximo passo: prática em R

```r
library(tidyverse)

dados <- read_csv("Advertising.csv")

dados %>%
  ggplot(aes(x=TV, y=sales)) +
  geom_point(col="darkred")+
  labs(y="Sales", x="TV")

dados %>%
  ggplot(aes(x=radio, y=sales)) +
  geom_point(col="darkred")+
  labs(y="Sales", x="Radio")

dados %>%
  ggplot(aes(x=newspaper, y=sales)) +
  geom_point(col="darkred")+
  labs(y="Sales", x="Newspaper")

fit_lm <- lm(sales ~ ., data=dados)

pred <- predict(fit_lm)
```

Avaliando a qualidade do ajuste

- Precisamos de uma maneira de medir a performance de um método de aprendizagem estatística.
- Quão boas as predições são em relação aos dados observados?
- No contexto de regressão, a medida mais utilizada é o *erro quadrático médio* (MSE - *mean squared error*):

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{f}(x_i))^2.$$

- $\hat{f}(x_i)$ é a predição que $\hat{f}$ dá para a i-ésima observação.
- *MSE* vai ser pequeno se as predições forem próximas dos valores verdadeiros.
- **MSE de treino**: calculado com as observações do conjunto de treino/treinamento.

Valor esperado

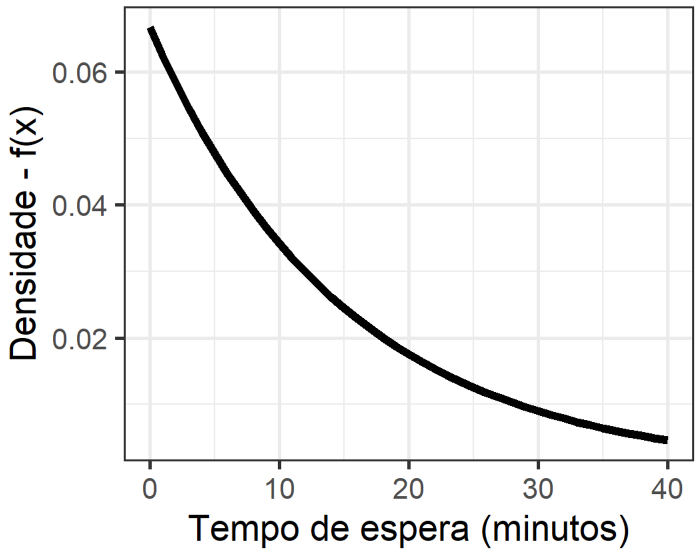
- Variável aleatória discreta

prêmio	P(prêmio)
50	0,7
100	0,2
200	0,1

$$E(\text{premio}) = 50 \times 0{,}7 + 100 \times 0{,}2 + 200 \times 0{,}1.$$

Valor esperado

- Variável aleatória contínua: o tempo de espera em uma estação de metrô segue uma distribuição exponencial com média de 15 minutos. $X \sim \text{Exp}(\lambda = 1/15)$.



$$E(X) = \int_0^\infty x f(x) dx = 15.$$

Erro de predição esperado

- Em termos matemáticos, o erro de predição esperado é dado por

$$E[(Y - f(x))^2].$$

- Objetivo: encontrar uma função $g : \mathbb{R}^p \to \mathbb{R}$ que minimize o erro de predição.
- Suponha que $(X, Y) \sim F_{X,Y}$.
- $E[(Y - f(x))^2] = \int (y - f(x))^2 dF_{X,Y}(x, y)$.
- Utilizando as propriedades da esperança condicional, podemos reescrever a expressão acima

$$\begin{aligned} E[(Y - f(x))^2] &= E[E[(Y - f(x)^2)|X]] \\ &= \int E[(Y - f(x))^2 | X = x] dF_X(x). \end{aligned}$$

- Então, basta encontrar f que minimize o integrando $E[(Y - f(x))^2 | X = x]$.

- **Minimizar** $E[(Y - f(x))^2|X = x]$.

- Desenvolvendo o produto notável do tipo $(a + b)^2$, temos:

$$\begin{aligned} E[(Y - f(x)^2)|X = x] &= E[Y^2 - 2Yf(x) + f^2(x)|X = x] \\ &= E[Y|X = x] - 2f(x)E[Y|X = x] + f^2(x). \end{aligned}$$

- Derivando em relação à $f(x)$ e igualando a zero:

$$-2E[Y|X = x] + 2f(x) = 0.$$

- Logo, a função f que minimiza o erro de predição esperado é a função de regressão

$$f(x) = E[Y|X = x].$$

- Problema?

- Na prática, trabalhamos com amostras e não sabemos a distribuição $F_{X,Y}$!

E agora?

- A solução de $E[Y|X = x]$ depende de quantidades desconhecidas, $F_{X,Y}$.
- Entretanto, uma boa estimativa para $E[Y|X = x]$ nos fornecerá um bom preditor para Y.
- Considere uma amostra $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ dos dados.
- **Ideia central da aprendizagem estatística:** utilizar a informação contida na amostra para construir uma função $\hat{f} : \mathbb{R}^p \to \mathbb{R}$ que será a estimativa para f.

Revisando

- Lembre-se que estamos considerando o modelo aditivo

$$Y = f(X) + \epsilon,$$

em que X é o vetor de preditoras, com distribuição F_X e o erro aleatório ϵ tem média 0, variância σ^2 e é independente de X.

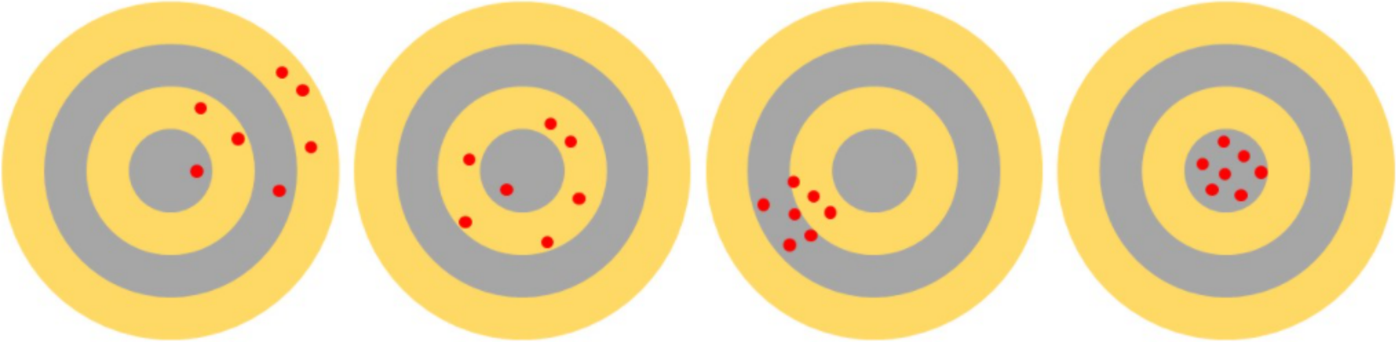
- A utilização do modelo aditivo nos permite decompor o erro de predição esperado $E[(Y - \hat{f}(X))^2]$ em fatores interpretáveis: **erro redutível** e **erro irredutível**.

- Além disso, ele pode ser escrito como

$$\begin{aligned} E[(Y - \hat{f}(x))^2] &= E[E[(Y - \hat{f}(x))^2 | X]] \\ &= \int E[(Y - \hat{f}(x))^2 | X = x] dF_X(x). \end{aligned}$$

Relação entre viés e variância

- $\text{Var}(\theta) = E(\theta^2) - E^2(\theta)$;
- $\text{Viés}(\theta) = E(\hat{\theta}) - \theta$;
- $\text{MSE}(\hat{\theta}) = E[(\hat{\theta} - \theta)^2] = \text{Viés}^2(\theta) + \text{Var}(\hat{\theta})$.



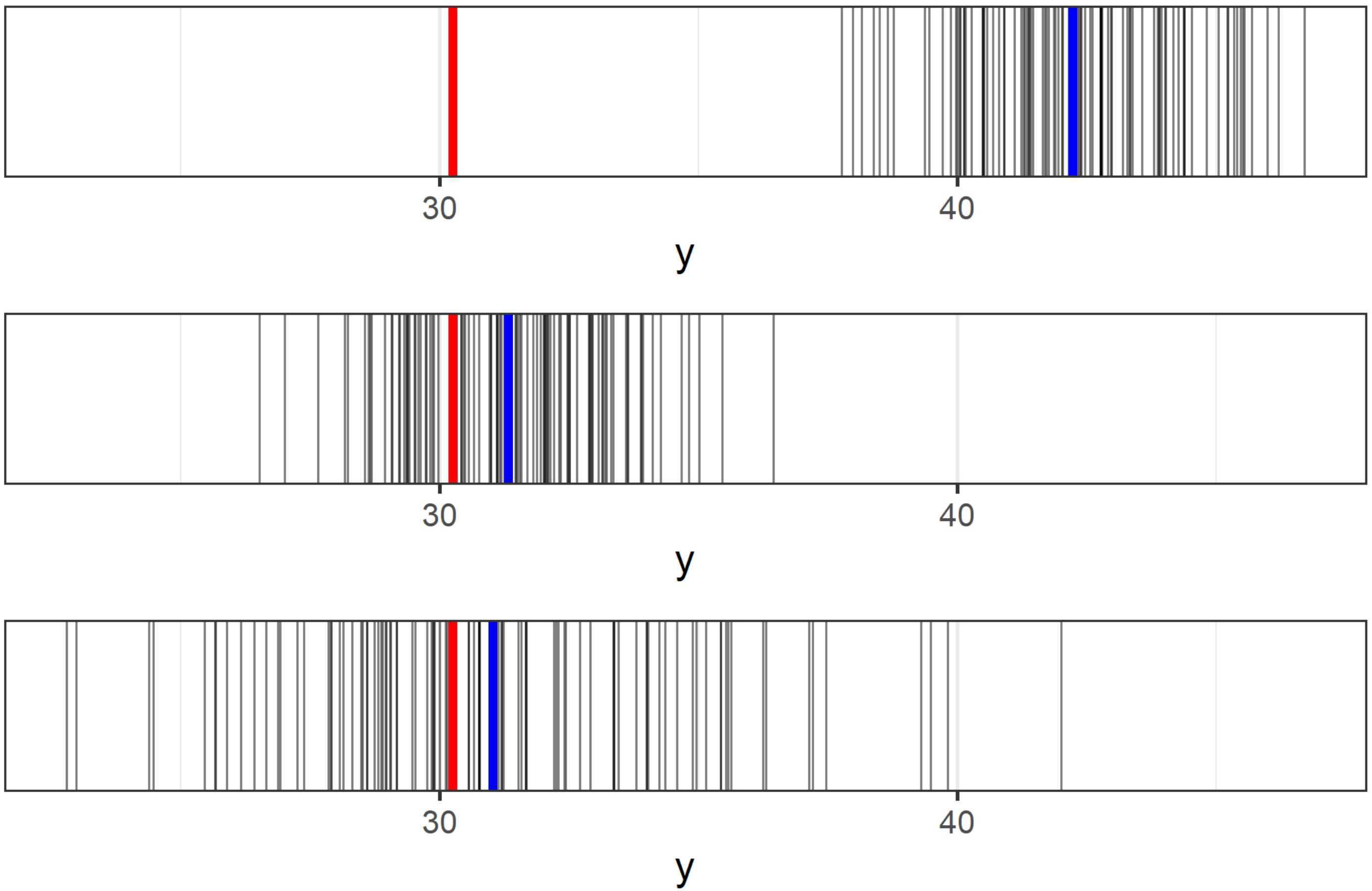
Decomposição do erro em viés e variância

- Podemos decompor $E[(Y - \hat{f}(X))^2]$ em termos de viés e variância:

$$
\begin{aligned}
E[(Y - \hat{f}(x))^2] &= \int (\text{Vies}^2(\hat{f}(x)) + \text{Var}(\hat{f}(x)))dF_X(x) + \sigma^2 \\
&= \int \text{MSE}[\hat{f}(x)]dF_X(x) + \sigma^2.
\end{aligned}
$$

- O resultado acima nos diz que para minimizar o erro de predição esperado, temos que selecionar um método de aprendizagem estatística que tenha baixo viés e baixa variância simultaneamente.

Viés e variância em três cenários



Trade-off entre viés e variância

- Dados de treinamento são usados para estimar f.
- Diferentes conjuntos de dados de treinamento geram $\hat{f}$ distintas.
- Idealmente, $\hat{f}$ não deve variar tanto ao usar um conjunto de treinamento diferente.
- Decomposição do erro:

$$\mathrm{MSE}[\hat{f}(x)] = \mathrm{Vies}^2[\hat{f}(x)] + \mathrm{Var}[\hat{f}(x)]$$

- A **variância** de um método de aprendizagem estatística mede o quanto a estimativa $\hat{f}$ muda conforme treinamos o modelo com novos dados de treinamento.
- O **viés** de um método de aprendizagem mede o quanto $\hat{f}$ difere em média de f sob replicação do processo de aprendizagem.
- Em geral, modelos mais flexíveis apresentam alta variância e baixo viés.
- **Trade-off**: é fácil diminuir o viés do aprendizado aumentando a sua variância (e vice-versa).

Simulação

- Estamos interessados em prever a renda anual Y de uma pessoa que possui x anos de escolaridade.

- Considere que a função que relaciona anos de escolaridade com a renda anual é

$$Y = f(x) = 45 \times \tanh\left(\frac{x}{1{,}9} - 7\right) + 57.$$

Geração de dados

- Vamos considerar pessoas com anos de estudo distribuídos uniformemente entre 8 e 18, ou seja, $X \sim U(8, 18)$.

- Para simular a realidade, adicionamos um erro:

$$Y = f(x) + \sigma^2.$$

- O erro aleatório σ^2 segue uma distribuição normal com média 0 e desvio padrão 4, ou seja, $\sigma^2 \sim N(0, 4^2)$.

- Objetivo: avaliar o modelo para previsão Y quando $x = 10$.

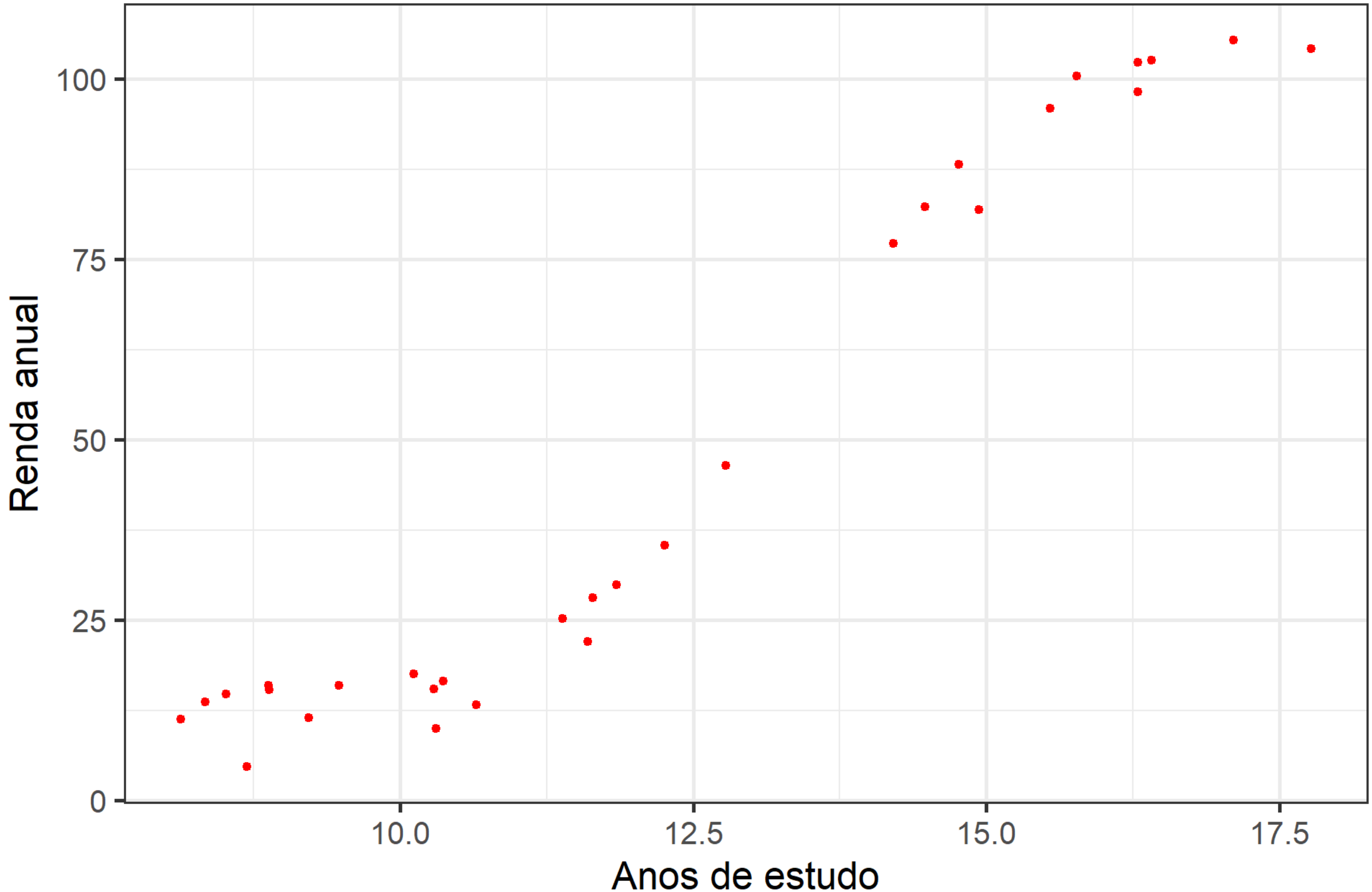
- $f(10) = 14, 7$.

Como fazer no R?

```
n_obs <- 30

valor_x <- runif(n = n_obs, min = 8, max = 18)

valor_y <- 45*tanh(valor_x/1.9 - 7) + 57 + rnorm(n = n_obs, mean =

dados <- tibble(x = valor_x,
                y = valor_y)

dados %>%
  ggplot(aes(x, y)) +
  geom_point()
```

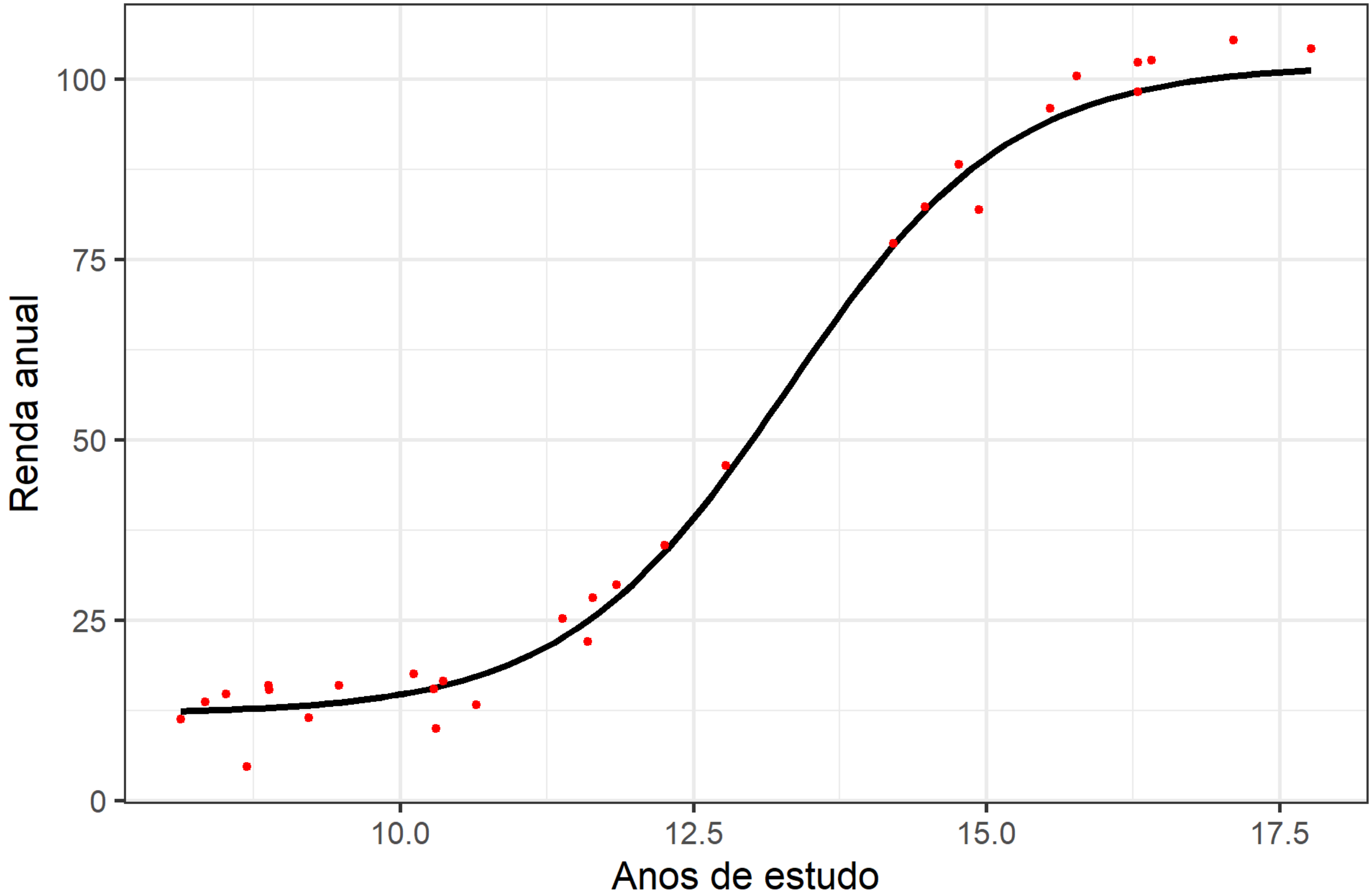
Dados gerados

Veja abaixo uma amostra de 30 "pessoas" com anos de estudo e salário anual (x_i, y_i) gerados de acordo com a definição anterior.

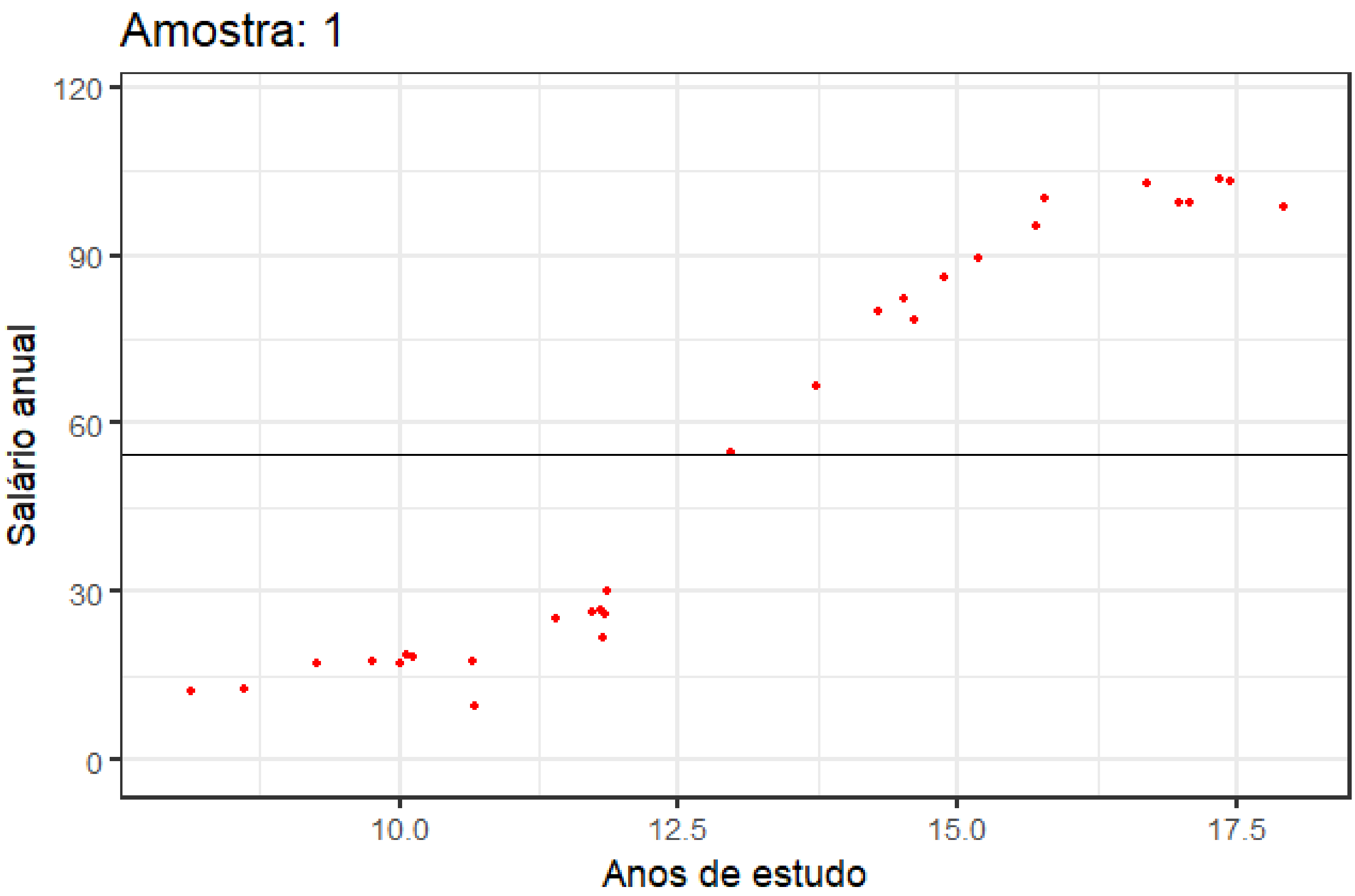


Dados gerados

Veja abaixo uma amostra de 30 "pessoas" com anos de estudo e salário anual (x_i, y_i) gerados de acordo com a definição anterior.

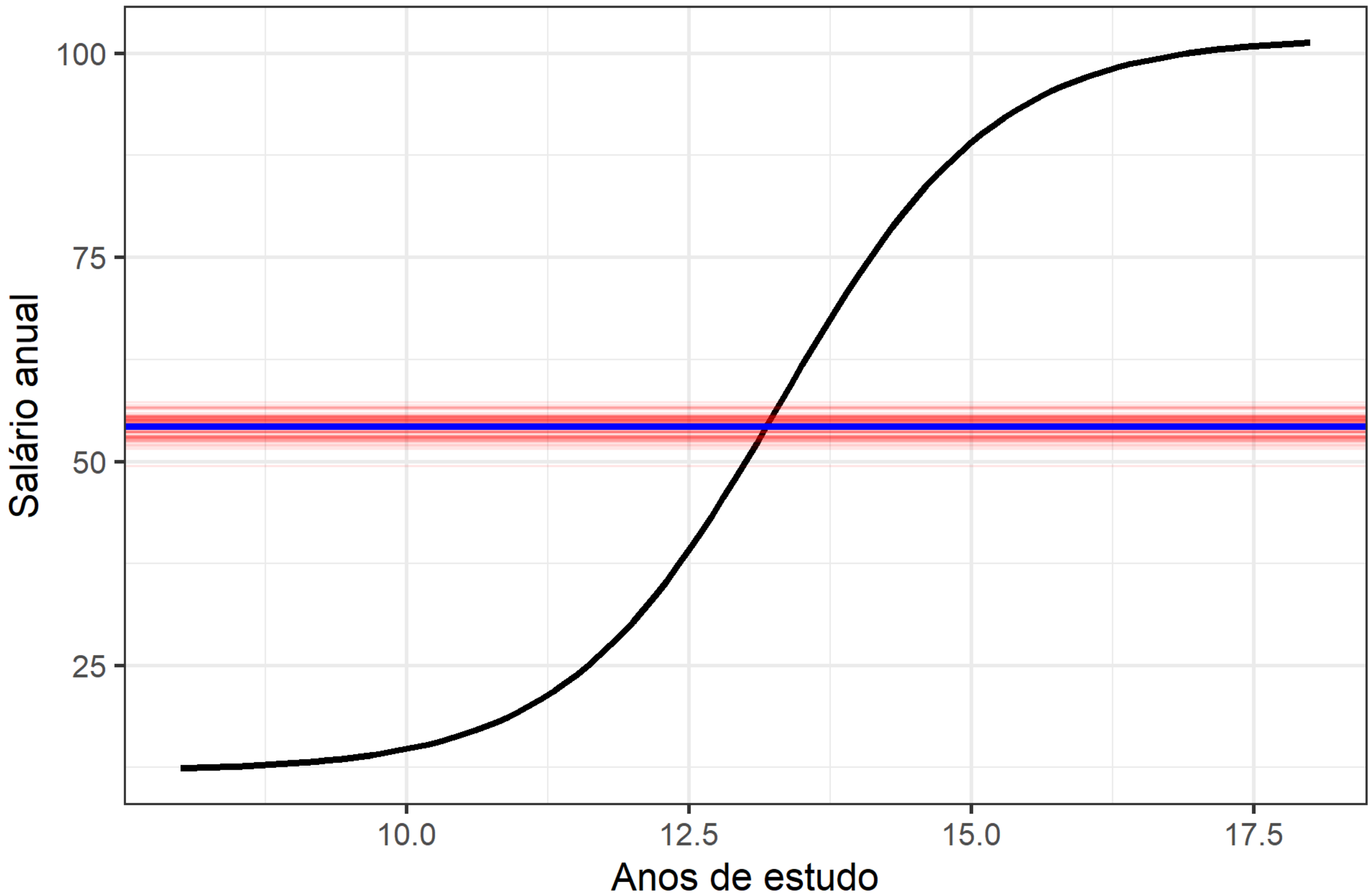


Várias amostras



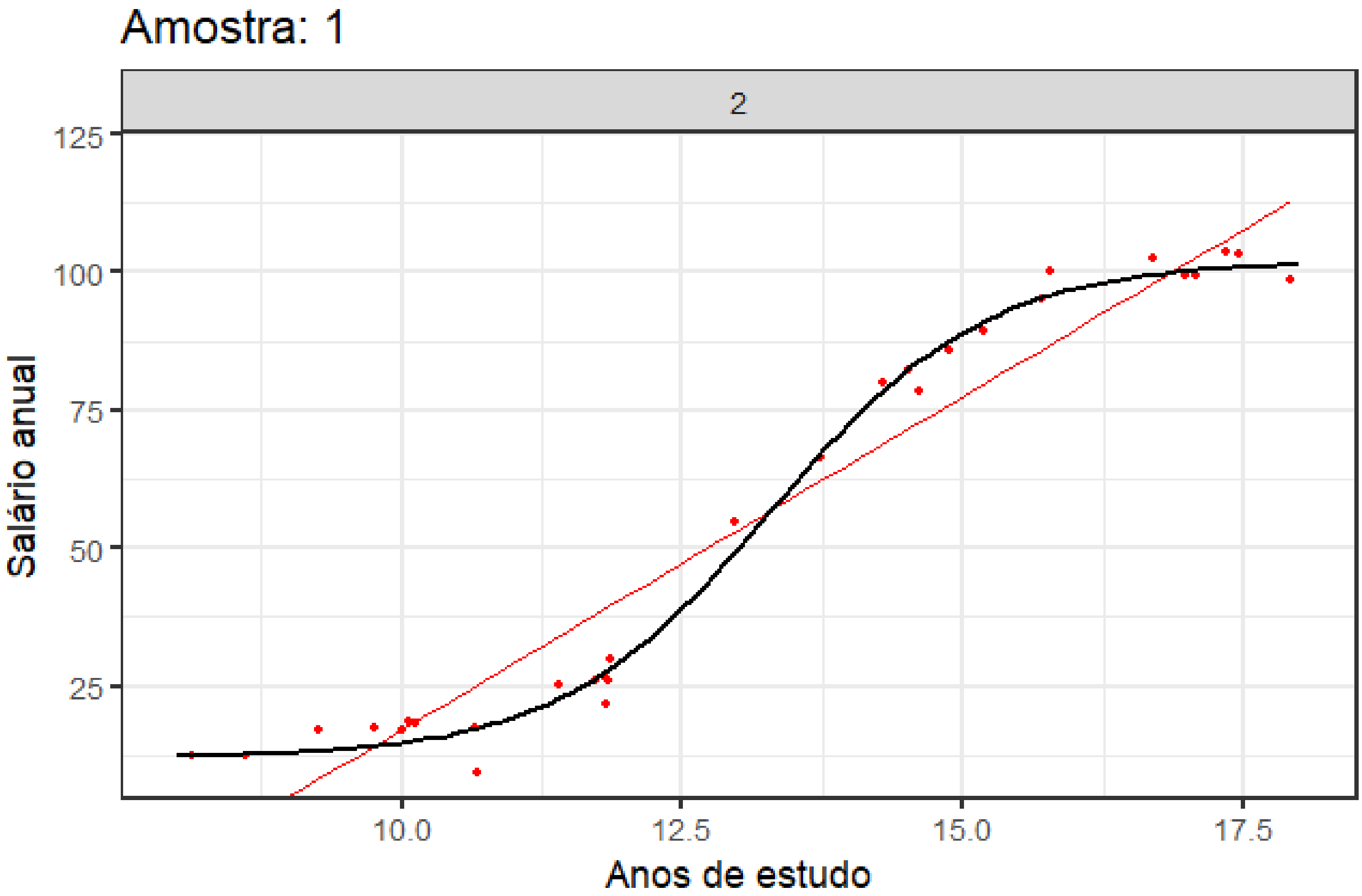
Modelo médio

Considerando os 100 conjuntos de dados gerados e $\hat{f}(x) = \frac{1}{30}\sum_{i=1}^{30} Y_i$.



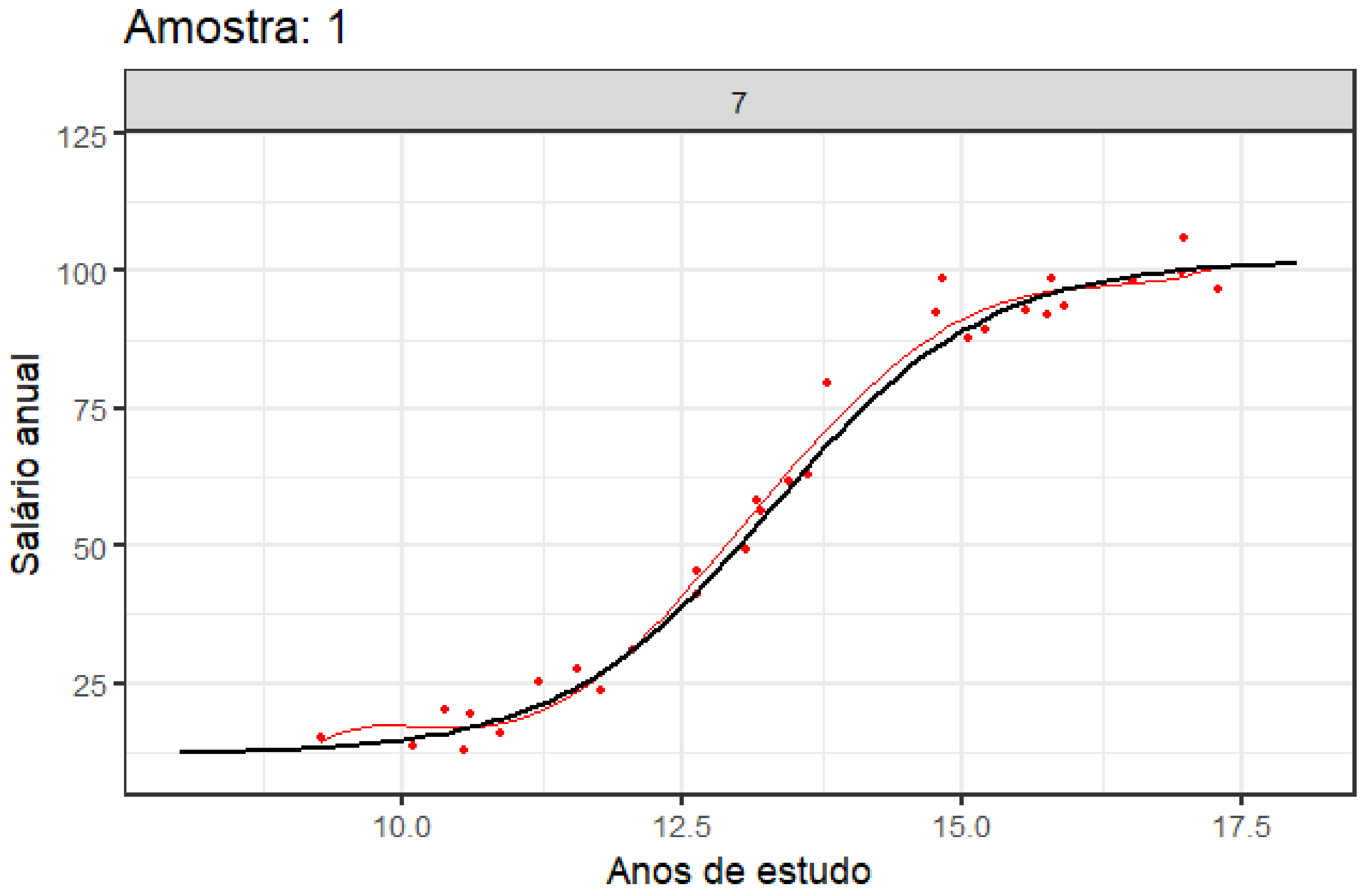
Modelos mais flexíveis

Considerando 100 modelos com 2 graus de liberdade.



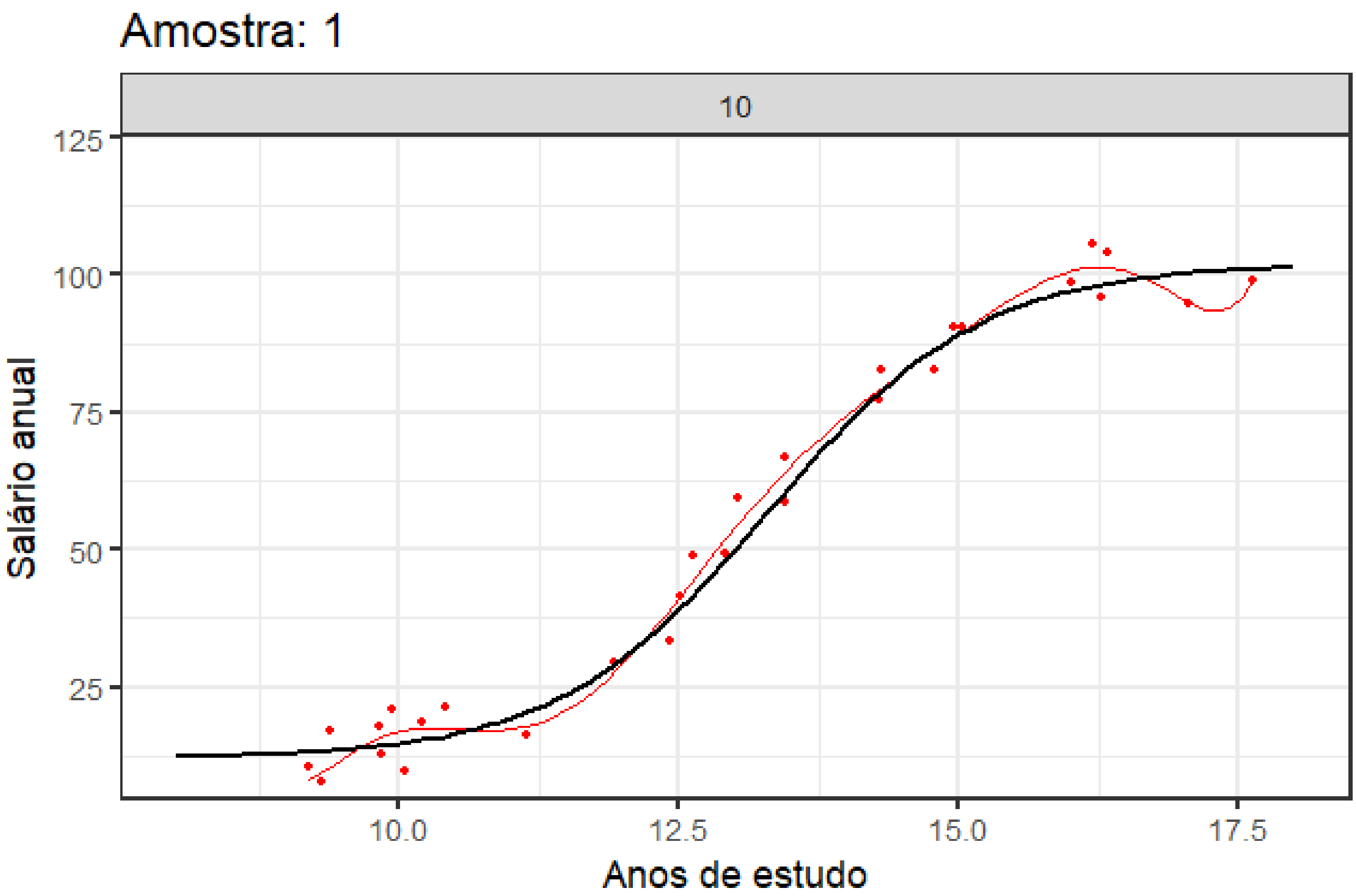
Modelos mais flexíveis

Considerando 100 modelos com 7 graus de liberdade.



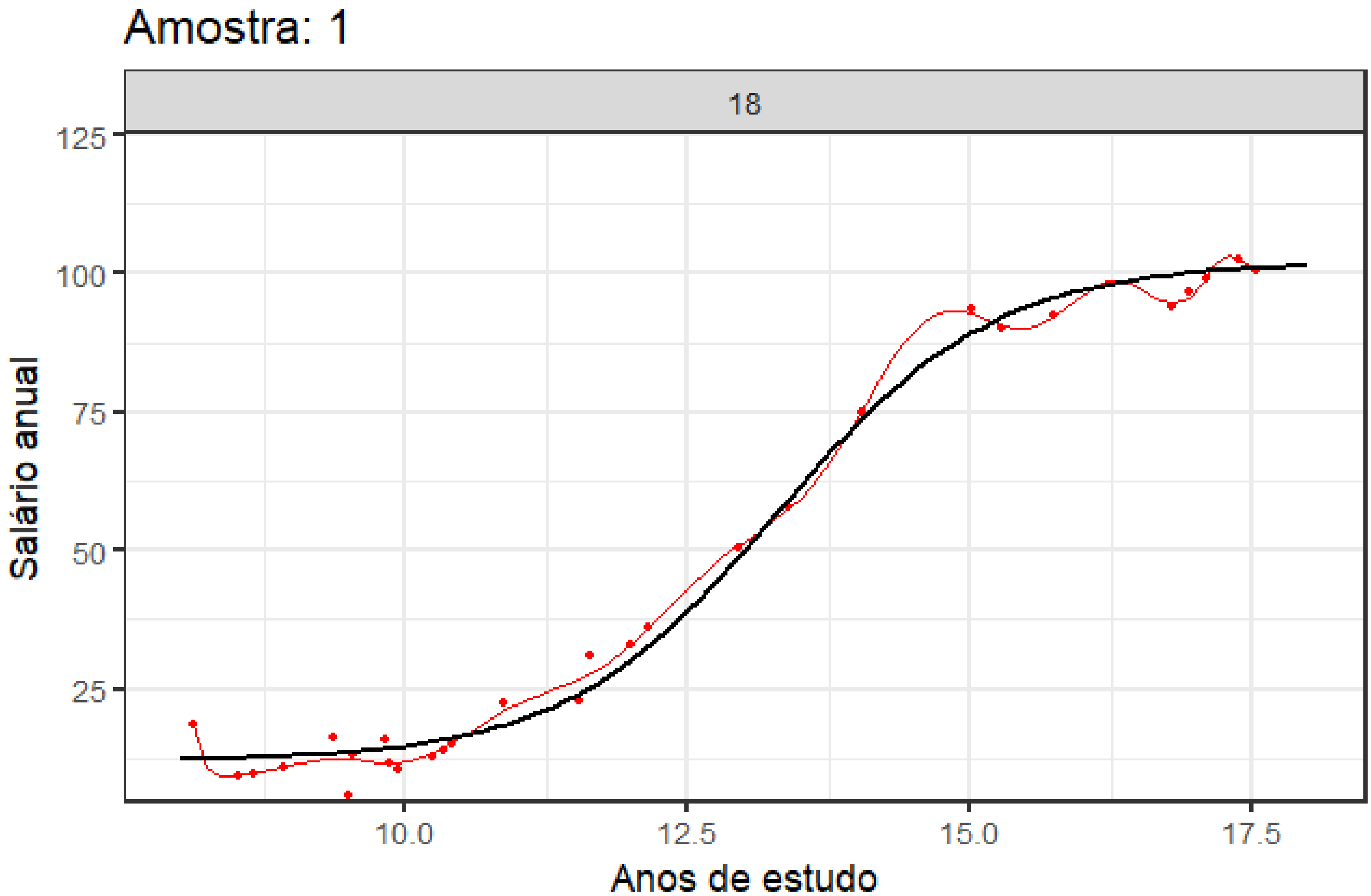
Modelos mais flexíveis

Considerando 100 modelos com 10 graus de liberdade.

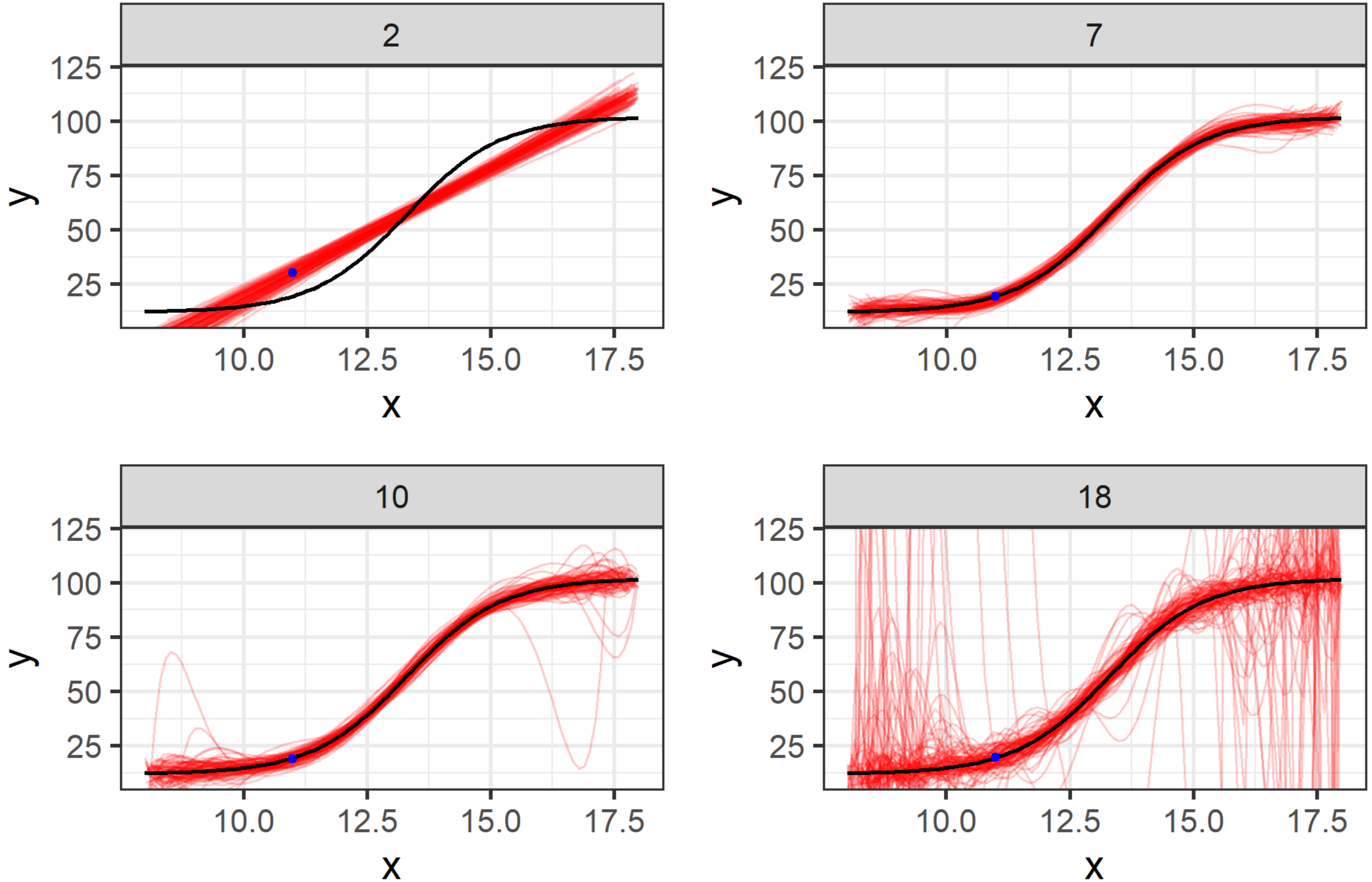


Modelos mais flexíveis

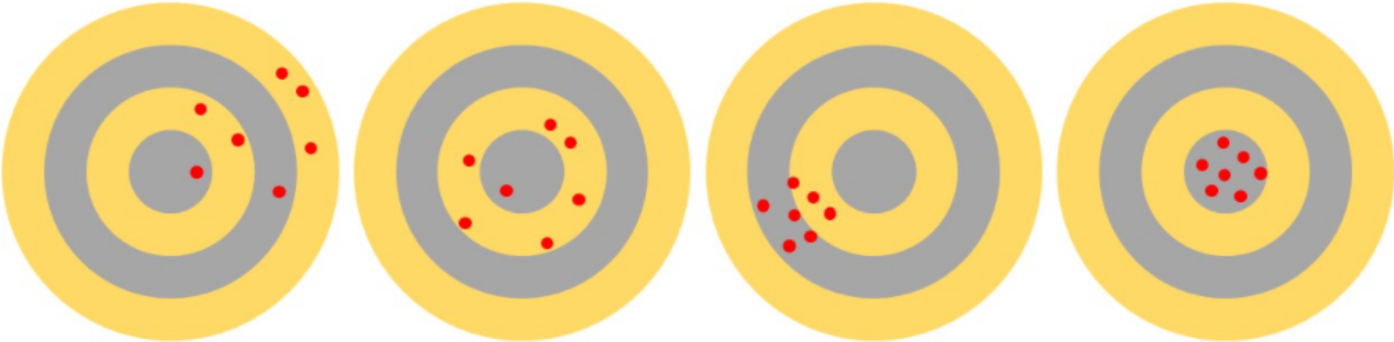
Considerando 100 modelos com 18 graus de liberdade.



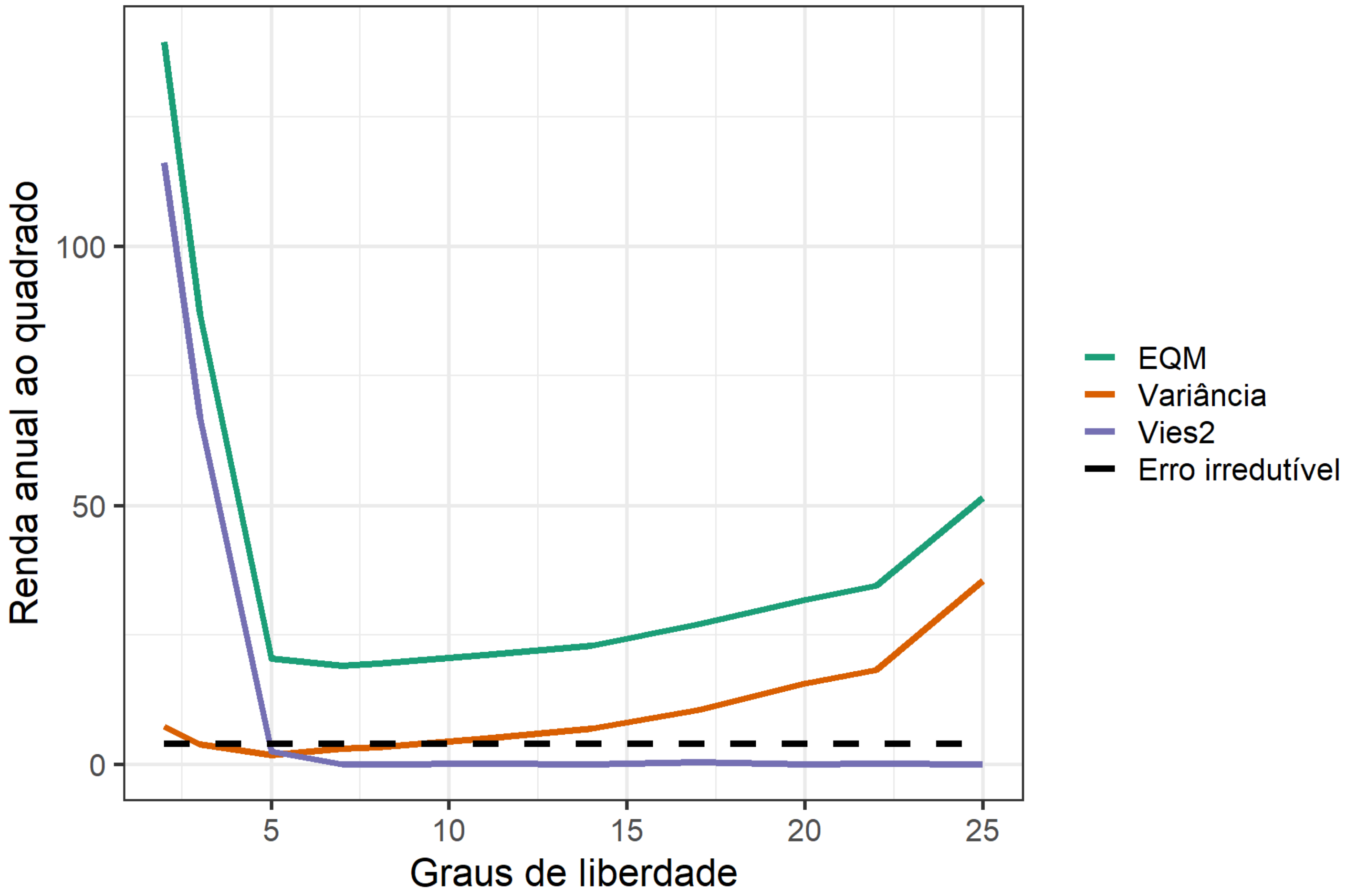
Modelos mais flexíveis



Qual a relação desta figura com o slide anterior?



Trade-off viés e variância

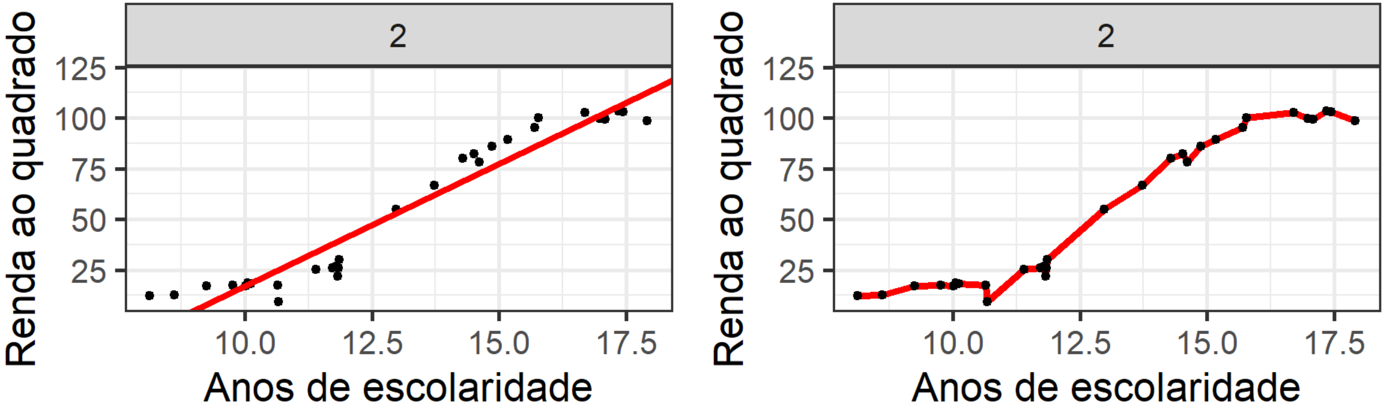


Problema

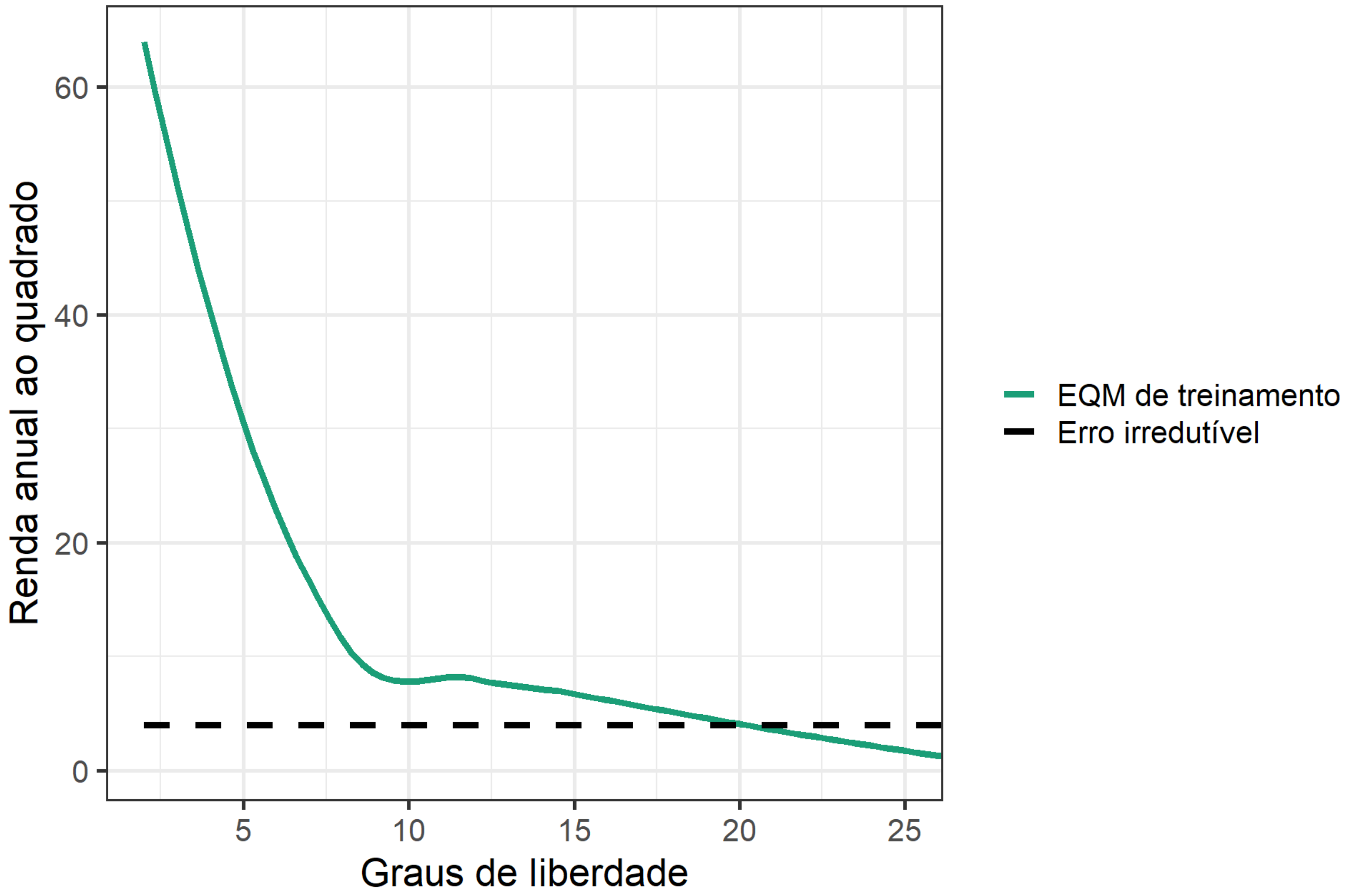
- Na prática, não conhecemos a função $f(\cdot)$!
- Como estimar o erro de predição $(f(x_i) - \hat{f}(x_i))^2$?
- Alternativa: considerar a diferença $(y_i - \hat{f}(x_i))^2$

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{f}(x_i))^2.$$

- Mas existe um problema!



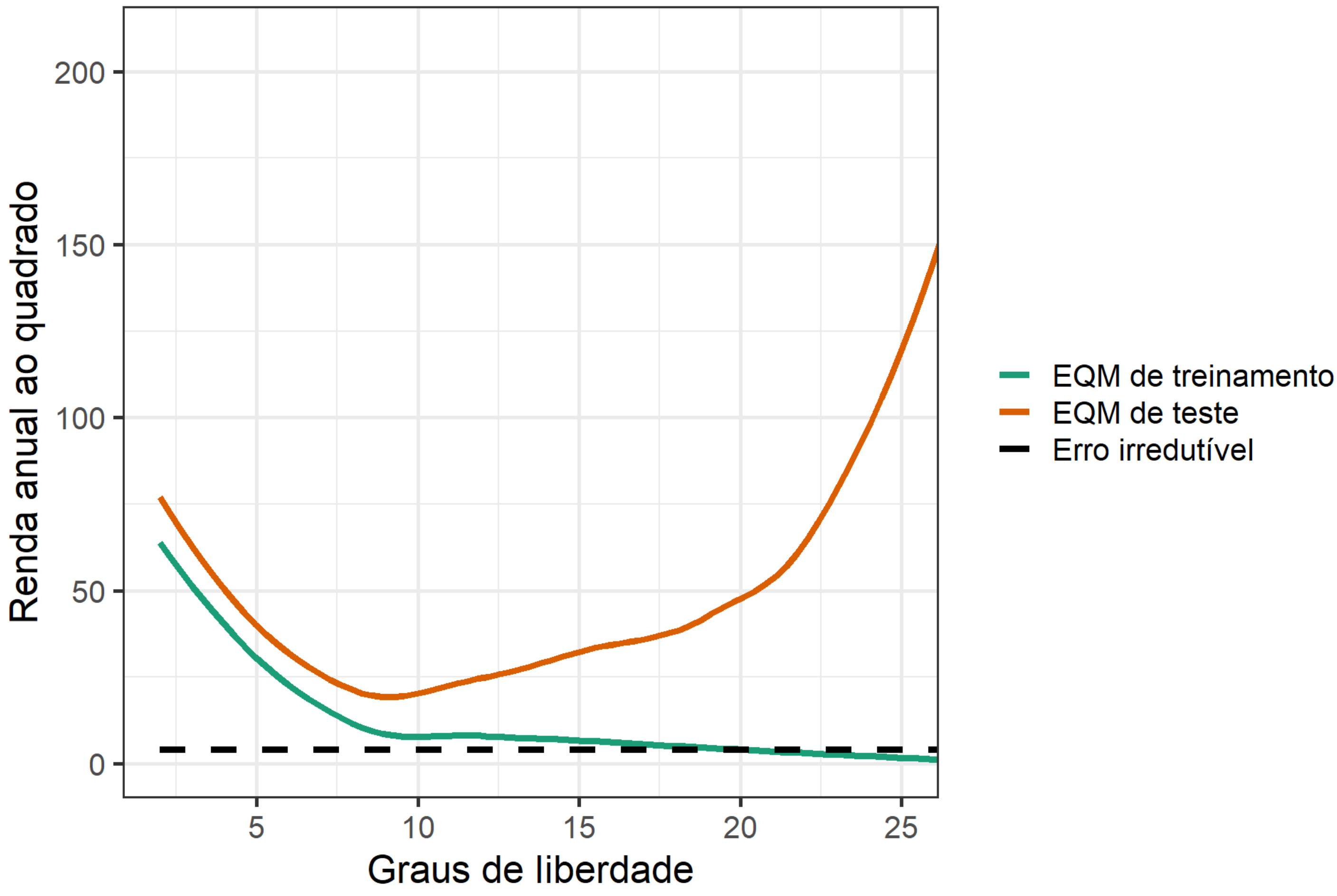
Erro quadrático médio de treinamento



O que fazer se não houver conjunto de teste?

- **MSE de teste**: calculado com dados que **não** pertencem ao conjunto de treinamento.
- Essa métrica é utilizada na seleção de modelos.
- Na prática não há um conjunto de teste disponível.
- Alternativas:
  - *Validation set approach*;
  - *Cross validation*;
  - *Bootstrap*;
  - $\rightarrow$ assunto das próximas aulas!

Erro quadrático médio de treinamento e de teste



Resumindo...

- Definição de variável resposta Y e variáveis preditoras $X = (X_1, \dots, X_p)$.
- Quando Y é quantitativa, dizemos estar tratando de um problema de **regressão**.
- Modelo: $Y = f(X) + \epsilon$, f é desconhecida.
- Queremos estimar f para fazer **inferência** ou **predição**.
- Estimativa de f: $\hat{f}$.
- O erro de $\hat{f}$ pode ser decomposto em
  - erro **redutível**: $[f(X) - \hat{f}(X)]^2$,
  - erro **irredutível**: $\mathrm{Var}(\epsilon)$.
- Relação entre interpretabilidade e flexibilidade.
- Qualidade de um ajuste pode ser avaliada através do $\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{f}(x_i))^2$.
- *Trade-off* entre viés e variância: $\mathrm{MSE}[\hat{f}(x)] = \mathrm{Vies}^2[\hat{f}(x)] + \mathrm{Var}[\hat{f}(x)]$.
- Performance do modelo deve ser avaliada no conjunto de **teste**.

Referências

An Introduction to Statistical Learning: with Applications in R. James, G. and Witten, D. and Hastie, T. and Tibshirani, R. 2013.

Notas de aulas prof. Tiago Mendonça. Curso PADS Modelos Preditivos.

Obrigado!

magnotfs@insper.edu.br